A minimal dynamical model of intonation

Citation for published version:
Iskarous, K, Cole, J & Steffman, J 2024, 'A minimal dynamical model of intonation: Tone contrast, alignment, and scaling of American English pitch accents as emergent properties', *Journal of Phonetics*, vol. 104, 101309. https://doi.org/10.1016/j.wocn.2024.101309

Digital Object Identifier (DOI):
10.1016/j.wocn.2024.101309

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Phonetics

A Minimal Dynamical Model of Intonation: Tone contrast, alignment, and scaling of American English pitch accents as emergent properties

Corresponding Author: Khalil Iskarous
University of Southern California
kiskarou@usc.edu

Jennifer Cole
Northwestern University
jennifer.cole1@northwestern.edu

Jeremy Steffman
The University of Edinburgh
jeremy.steffman@ed.ac.uk

Abstract

The pitch accent system of Mainstream American English (MAE) is one of the most well-studied phenomena within the Autosegmental-Metrical (AM) approach to intonation. In this work we present an explicit model grounded in dynamical theory that predicts both qualitative phonological and quantitative phonetic generalizations about the MAE system. While the traditional AM account separates a phonological model of the structure of the accents from the F0 algorithm that interprets the phonological specification, we propose a unified dynamical model that encompasses both. The proposed model is introduced incrementally, one dynamical term at a time, to arrive at the minimal model needed to account for observed empirical generalizations, avoiding unnecessary complexity. The quantitative and qualitative properties of the MAE system that inform the dynamical model are based on an analysis of a large database of productions of the four most well-studied pitch accents of American English: three rising accents (H*, L+H*, L*+H) and a low-falling accent (L*). The dynamic model highlights the importance of velocity-based measures of F0, not typically invoked in intonational research, as key to understanding F0 differences among pitch accent categories. Although the focus of this work is on the MAE pitch accent system, suggestions are made for how the unified phonetic-phonological dynamical framework presented can be further developed to account for other pitch-based phenomena in a variety of languages.

1 Introduction

The Autosegmental-Metrical (AM) approach to intonation (Pierrehumbert, 1980; Ladd, 1996/2008; Beckman et al., 2005; Arvaniti, 2022) has been successful at describing a variety of pitch accent systems in the world's languages (Jun 2005, 2014). The empirical reach of AM is especially admirable given the small number of constructs assumed, e.g., the phonological level tone features (H and L), the cross-tier alignment operator (*), the linear sequence operator (+), and the notions of pitch range or scale. The purpose of this work is to propose a dynamical systems theory of pitch accent from which the constructs listed above *emerge*. Furthermore, this theory, as an instance of other dynamical approaches to speech (Fowler et al., 1980; Saltzman and Munhall, 1989; Browman and Goldstein, 1989; Byrd and Krivokapić, 2021; Iskarous and Pouplier, 2022), integrates the phonological and phonetic aspects of pitch organically, so there is no necessity for an F0 algorithm that interprets the phonological constructs above (cf. Pierrehumbert, 1981). A dynamical theory describes an invariant functional relationship between the current value of a variable, e.g., F0, and the change or velocity of that variable over successive moments in time. Even though F0 and its velocity may each vary continuously throughout a pitch accent, in the dynamical model presented here their relation is quite abstract and constant for some phonologically defined duration. We will show that the relation between the value of F0 at any time point during the interval of the accent and its velocity, or slope, is the result of a balance of *self-organizing* dynamical forces, whose tuning results in the inventory of a pitch accent system. The fact that dynamical system descriptions simultaneously account for continuous phonetic variation and abstract phonological invariance makes them the right kind of theory from which to derive Autosegmental-Metrical (AM) constructs, as we will argue. Specifically, we will show that these constructs, including the observed contrasts between level tones, variation in the temporal alignment of F0 target values, the relation between alignment and F0 scaling, along with shape distinctions among F0 trajectories, are predicted from the proposed dynamical system, arising from changes in the balance of dynamic forces that shape the F0 trajectory, tuned by a single parameter k. One additional force-balance parameter b, closely related to k, widens and narrows the scale, i.e., local pitch range, through interaction with k. The proposed dynamical model captures many properties of the AM model for American English, and it also goes beyond the AM model in accounting for certain shape-based properties of F0 trajectories that are observed in empirical data but which are not predicted by AM's target-and-interpolation approach. Moreover, this work aims to upgrade dynamical theories of speech phenomena by making them more emergentist and less stipulative, in an attempt to resolve a fundamental problem in task dynamical approaches pointed out by Pierrehumbert and Pierrehumbert (1990).

Despite its success, there have been a variety of critiques brought up against the AM theory and its level-based precursors. The first and most notorious one is the critique of the notion of *level* (Bolinger 1951) and advocacy for *configuration* approaches based on the shape of F0 contours (Fujisaki 1997; Hirst & di Cristo 2000) or for the inclusion of shape parameters in a level-based model (Barnes, Veilleux, Brugos & Shattuck-Hufnagel, 2012; Niebuhr 2007; Niebuhr, et al., 2011; D'Imperio 2000). The dynamical approach is antithetical to this long *levels-vs-configuration* debate, since a dynamical approach is about the abstract relationship between the state of a variable and changes in that state, so instead of further engagement in this debate, we aim to provide an approach that bypasses it. A second criticism is the difficulty of comparing the pitch accents of different languages and dialects (Ladd, 2008a; Arvaniti, 2019), since what may be described with one set of AM representational constructs, e.g., L+H* in English and Spanish (Hualde and Prieto, 2016) or L+H* in varieties of American English (Burdin, Holliday & Reed, 2022), can be quite different in their F0 trajectories. It will be shown in this work that variation in the k parameter, the main determinant of the pitch accent, leads not only to quantitatively different F0 trajectories, but also to *qualitatively* different trajectories. Access to quantitative and qualitative differentiation allows for a theory in which languages and language varieties differ in which values of k anchor distinctions in their intonational phonologies. This allows us to bypass the dual use of phonological and broad phonetic descriptors (Hualde and Prieto, 2016) to solve this problem (where e.g., two phonologies can have an L+H* pitch accent, with different broad phonetic descriptions).

A third major criticism of AM arises in the analysis of languages like English and German, in which the intonation system include an inventory of pitch accent categories that differ in their tonal specification, yet where the implementation of those accents displays substantial variability in F0 trajectories, resulting in overlap among distinct accent categories (Arvaniti 2016, 2019; Cangemi & Grice, 2016; Cole, Steffman, Shattuck-Hufnagel &Tilsen, 2023; Grice et al., 2017). Phonetic variation of this sort is problematic for dualist theories like AM, where discrete phonological categories should map onto distributions along one or more phonetic parameters such that the distributions are differentiable. Variable F0 trajectories in the implementation of an accent does not pose the same challenge for dynamical system models. In this paper, after first introducing a deterministic dynamical system model of pitch accents, we show that the deterministic model is only a special case of more general stochastic systems. Where deterministic dynamical systems are defined in terms of states and how they change across increments of time, stochastic systems are defined in terms of distributions over states, and the evolution of those distributions over time. We argue that this stochastic property provides a natural model of phonetic variation in the production of pitch accents across instances and speakers.

The theory we present will be tested based on an extensive empirical database of Mainstream American English (MAE) pitch accents, produced by 130 speakers (Cole et al., 2023; Steffman, Shattuck-Hufnagel & Cole, 2023). However, in the exposition of the theory and its discussion we will also present how the force-balance system built for MAE can, in future work, be configured to account for a variety of known pitch accent phenomena in other languages. Our approach is to develop the dynamical system incrementally, showing how simple dynamical system modules can be connected to yield different kinds of complex pitch behaviors. The goal is to provide a modular parametric theory that can capture observed patterns of quantitative and qualitative variation in the F0 trajectories of accents in MAE, and across dialects and languages. This is possible because, even though dynamical systems operate over real-valued variables, here F0, the dynamical relations between the variable and its change over time predict stable levels, thresholds, and ranges, making dynamical systems theory appropriate for deriving phonological constructs necessary for the descriptions of dialectal and linguistic variation.

We begin in Section 2 with a brief overview of theoretical foundations in earlier work on dynamical approaches to the analysis of phonological and speech motor systems. Section 3 introduces the empirical data which informs the development of our dynamical system, focusing on the F0 measures that capture essential properties of the four MAE pitch accents we focus on , discussed in terms of dynamical properties and the representational devices of the AM model. Section 4 presents the dynamical system, starting from the simplest dynamical system all the way to the minimal dynamical system that can account for the generalizations. Section 5 discusses the theoretical contributions of this work, its limitations, and future directions.

2. Theoretical foundations

Our approach builds on methods used first by Goldsmith, and later others, who derive discrete constructs of phonology related to syllables and stress through models of continuous spreading-activation computation (Goldsmith and Larson, 1990; Goldsmith, 1994; Prince, 1993; Iskarous and Goldstein, 2018). Phonological structure, in this view, emerges from interactive computation among adjacent representational units (e.g., segments or gestures). For syllabification, Goldsmith and Larson (1990) show that syllable-defining sonority waves emerge from excitatory and inhibitory interaction between contiguous segments, with each segment influencing the sonority value of its neighbors, giving rise to complex syllabic phenomena, e.g., as seen in Tamazight Berber and Indonesian. For quantity insensitive stress systems, Goldsmith (1994) shows that properties such as boundedness, rhythmic alternation, and extrametricality can arise through continuous excitatory and inhibitory interactions between contiguous syllable prominences. Further, Prince (1993) shows that phonological properties like the limitation of stress to the periphery of a domain (e.g., 3 syllables from the end), what he calls a *barrier*, emerges from continuous-time scalar computations as proposed by Goldsmith. Prince (2005) lauds this approach to phonology as being *free-standing* in that "many predictions and properties can be determined from examination of the theory alone" (Prince, 2005). The predictive power of these computational theories stems from the simplicity and modularity of the fundamental computation, from which representational constructs are derived, as opposed to being posited at the basis of theory.

Our proposal also builds on work over several decades investigating speech motor control, perception, and phonological cognition from a dynamical perspective (Fowler et al., 1980; Browman and Goldstein, 1989; Saltzman and Munhall, 1989). We seek to *upgrade* this dynamical approach by showing how discrete phonological categories can *emerge* from continuous computation. Our upgrade is inspired by Pierrehumbert and Pierrehumbert's critique of Task Dynamics/Articulatory Phonology (AP) (Browman and Goldstein, 1990; Saltzman and Munhall, 1989) in their (1990) paper "On attributing grammars to dynamical systems". Pierrehumbert and Pierrehumbert (1990) argue that AP does not actually meet its stated goal of fully integrating the concrete/continuous and abstract/discrete aspects of language (Browman and Goldsmith, 1990). This can only be accomplished, Pierrehumbert and Pierrehumbert argue, if discrete cognitive aspects are derived from continuous dynamical computation. They correctly point out that in AP, the discrete is *specified*, not *derived.* In the words of the authors, "[t]hey [AP] do not fully bridge the gap between dynamics and phonology. 'Task dynamics' as it is most typically carried out to date takes *discrete* inputs and produces *continuously variable* outputs" (p.467). These "discrete inputs" in Task Dynamics includes dynamical parameters that specify the goal of a gesture, e.g., the degree of closure at a constriction, or specific place of articulation of a gesture. These are analogous to discrete featural specifications in traditional phonology.[1] Pierrehumbert and Pierrehumbert (1990) argue that in order to attribute grammars to dynamical systems, discrete equilibria at which a dynamical system settles and which are then associated with phonological contrasts (and, we believe, to syllable relations and prosodic units), should be *computed* or *derived*, not specified. We agree with this fundamental critique, but we do not believe that it is detrimental to dynamic theories. Qualitative constructs, like those of intonational phonology can indeed emerge, from dynamical theories if the parameters do not specify goals, but degrees of interaction between dynamical terms, as we will show in the rest of this paper. Indeed, Iskarous (2019) derives the very idea of a contrastive feature or gesture from dynamical computation. Our present work extends this approach for the analysis of intonation, and alongside Iskarous (2019), represents an attempt to advance contemporary phonological theory through emergentist computation.

The AM approach to intonation is subject to the same criticism that Pierrehumbert and Pierrehumbert (1990) level at AP in that the discrete constructs of level tones (L, H), tonal alignment (*) and tonal composition (+) are posited at the foundation of the theory, and continuous F0 trajectories are derived through an algorithm with discrete inputs (Pierrehumbert, 1980; Pierrehumbert, 1981; Beckman and Pierrehumbert; 1988; O'Shaugnessy, 1976; Maeda, 1976). In fact, this critique also applies to the dynamical systems model of F0 contours proposed by Yi Xu and colleagues (Xu et al. 1999; Prom-on et al. 2009, *inter alia*), which is a task-dynamic based linear dynamical system that models the combined change of F0, its velocity, acceleration, and jerk (time-change of acceleration). Their work is an important accomplishment that extends the results achieved in AP for supralaryngeal contrasts (Saltzman and Muhall, 1989; Browman and Goldstein, 1989) to laryngeal behavior and shows a good fit between predictions of the dynamical model with empirical F0 trajectories of Mandarin lexical tones and English stress/accent. However, their model, like AP, assumes the discrete equilibria that correspond to empirically observed articulatory or acoustic targets of phonological categories.

The dynamical models of F0 proposed by Xu and colleagues illustrate one additional property that we wish to challenge, and that is the reliance on second order and higher order linear models, meaning that the dynamical relations captured are between several derivatives of the main state variable (F0), its velocity, acceleration, and maybe jerk (the infinitesimal change in acceleration with time). We will argue that constructs central to the analysis of intonation can be meaningfully modelled using more conservative models in which acceleration (curvature of F0) or jerk play no role. Headway in this direction has already been made by Roessig, Mücke, and Grice (2019) and Roessig (2021) who introduce a dynamical model quite similar to the one we will present, though presented in that work as a model of the planning or selection of pitch accents in German[2], where a system equilibrium is associated with a

---

[1] More technically, AP includes a task equilibrium value within the task differential equation. We believe that this is true not only of AP of the time, but also extends to new versions of Task Dynamics (Byrd and Saltzman, 1998; Sorenson and Gafos, 2016), as well as modern extensions to syllable structure where equilibrium values of phase lags are pre-specified (Goldstein et al., 2007), and to boundary prosody where strength and specific phase-lag between articulatory gestures and the π-gesture is explicitly specified (Byrd and Saltzmann, 2003).

[2] The dynamical differential equation we present in Equation (1) can be obtained as the negative derivative of the potentials they provide.

speaker's choice of one accent vs. another. However, in this paper we show that that the shape of an F0 trajectory that corresponds to a pitch accent can also be modelled in terms of the same dynamic.

The ultimate goal in understanding prosody, of course, is to "establish a complete picture of a prosodic typology" (Jun, 2005). A full typology would allow for different aspects of prosody: pitch patterning, boundary-based segment/gesture duration, juncture type and duration, speech timing (rhythm), stress patterning, syllables, and tonal contrasts, to be stated in comparable terms not only for cross-linguistic description, but also for describing how any of these properties interact in a language when they co-occur. Since dynamical analyses of all these phenomena have been proposed (Gao, 2009; Hermes et al., 2013; Shaw and Gafos, 2015; Shaw et al., 2019; Iskarous and Goldstein, 2018; Katsika, 2016; Katsika et al., 2014; Karlin and Tilsen, 2014; Krivokapić , 2014, 2020; Krivokapić et al., 2020), we hope that the model we propose will in the future be able to interact with these other theories, to provide a common dynamical vocabulary for different aspects of prosodic systems, in addition to making within- and across-language differences clearer.

3. Methods

3.1 The empirical F0 trajectories

The pitch accent model we propose is based on data from a project led by two of us (Cole, Steffman) investigating the perception and production of MAE pitch accents in nuclear position (i.e., the final pitch accent in the prosodic phrase). The data are from 130 MAE speakers, aggregated from three imitative speech production experiments. In each experiment, on a given trial, participants heard model utterances that exemplified a particular tune. F0 in the model utterances was resynthesized based on straight-line approximations from the ToBI training materials (Veilleux, Shattuck-Hufnagel & Brugos, 2006), shown in Figure 1, which are in in turn based on empirical F0 trajectories in Pierrehumbert (1980). In all the experiments, the participant produced the heard tune on the final three-syllable, stress-initial name in a sentence that was metrically and syntactically similar to the stimuli (e.g., *heard*: "He answered Jeremy", "Her name is Marilyn"; *produced*: "They honored Melanie", with the target tune realized on the underlined material). Across all materials analyzed here, imitated tunes were produced on one of three names: Harmony, Madelyn, and Melanie.

The first experiment from which we drew data is described in Steffman, Shattuck-Hufnagel & Cole (2022). In that study, data from 70 speakers of MAE produced the high-toned pitch accents H*, L+H* and L*+H on a phrase-final word in all intonational boundary contexts—followed by one of four possible edge tone specifications: H-H%, H-L%, L-H%, L-L%. Each tonally unique sequence of pitch accent and edge tones was repeated 12 times for total of 3,360 productions per pitch accent (48 tokens of each pitch accent for 70 speakers). We also wanted to model productions of the low-toned L* pitch accent, which were lacking in that study, so we included data from two additional experiments. One data set is described in Cole, Steffman, Shattuck-Hufnagel & Tilsen (2023), in which 30 participants produced H* and L* pitch accents in all edge tone contexts. Given that we already had productions of H* from 70 other speakers, we opted to use just L*, with 72 L* tokens elicited per speaker, totaling 2,160. To include data from more speakers for the L* pitch accent, we added data from a third experiment. That experiment was a replication of Cole et al. (2023) that differed only in that participants heard two model sentences on each trial, as compared to the three model sentences in Cole et al. (2023), and they only produced two different sentences, as compared to three sentences as in Cole et al. (2023). This third experiment yielded another 2,160 tokens of L* for a combined total of 4,320 L* tokens aggregated from 60 speakers, with equal numbers in the four edge tone contexts. Because we expected phrase-final non-modal phonation to lead to potential issues in F0 measurement, F0 trajectories were audited for F0-tracking errors using a semi-automated procedure which made use of an algorithm detecting sudden F0 jumps (Steffman & Cole, 2022) and manual inspection of the files flagged by the algorithm. Approximately 10% of each data set was excluded due to inaccurate F0 measurement. This left 2,970 H*, 2,962 L+H*, 2,981 L*+H and 3,830 L* productions, produced by a total of 130 speakers. We measured F0 using STRAIGHT, as implemented in Voicesauce (Kawahara, Cheveign, Banno, Takahashi & Irino 2005; Shue, Keating, Vicenik & Yu, 2011).

As described above, our production data come from participants' imitations of resynthesized tunes presented as model utterances. Table 1 shows six F0 target values spanning each model speaker's pitch range, which defined the accentual and edge tone pitch targets of the resynthesized tunes for both the male and female model speakers. These six F0 targets correspond to the relative height of tonal targets of the pitch accent, phrase accent and boundary tone in each phrase-final intonational tune, based on published examples (Veilleux et al., 2006; Pierrehumbert, 1980), and are represented in the faint grid lines in the F0 trajectory plots in Figure 1. Figure 1 shows the time-normalized model tune trajectories in ERB, averaged across the two model speakers. Note that the values in ERB are centered to start at the value of zero for each trajectory. The reader is referred to Steffman, Shattuck-Hufnagel & Cole (2022) and Cole et al. (2023) for more details about the stimuli.

Note that the model tune F0 trajectories in Figure 1 span the entire nuclear accented word (e.g., *Melanie*), including not only the F0 trajectory corresponding to the pitch accent (the initial portion), but also the F0 trajectory representing the edge tones (the later portion). In this paper we have limited our focus to modeling the F0 trajectory of the pitch accent, in all four edge tone conditions. Accordingly, we segmented the full F0 trajectory of the imitated productions into two portions and used only the initial portion, corresponding to the region of the pitch accent, as the empirical basis for our dynamical systems model. This portion was the initial portion of the trajectory up to the F0 maximum (for H*, L+H*, L*+H) or minimum (L*) if that value occurred in the first two thirds of the trajectory, and otherwise the window for analysis ended at two thirds of the trajectory, which was the case for the monotonically rising or falling trajectories (e.g., H*H-H%, L*L-L%). This segmentation juncture is indicated for the model tune schema in Figure 1.

Table1: target values defined in each speaker's pitch range for the model stimuli.

	Male model speaker		**Female model speaker**	
	Hz	ERB	Hz	ERB
Target level 1	80	2.79	100	3.37
Target level 2	105	3.51	160	4.93
Target level 3	130	4.18	200	5.84
Target level 4	225	6.36	300	7.79
Target level 5	240	6.67	350	8.62
Target level 6	265	7.15	380	9.09

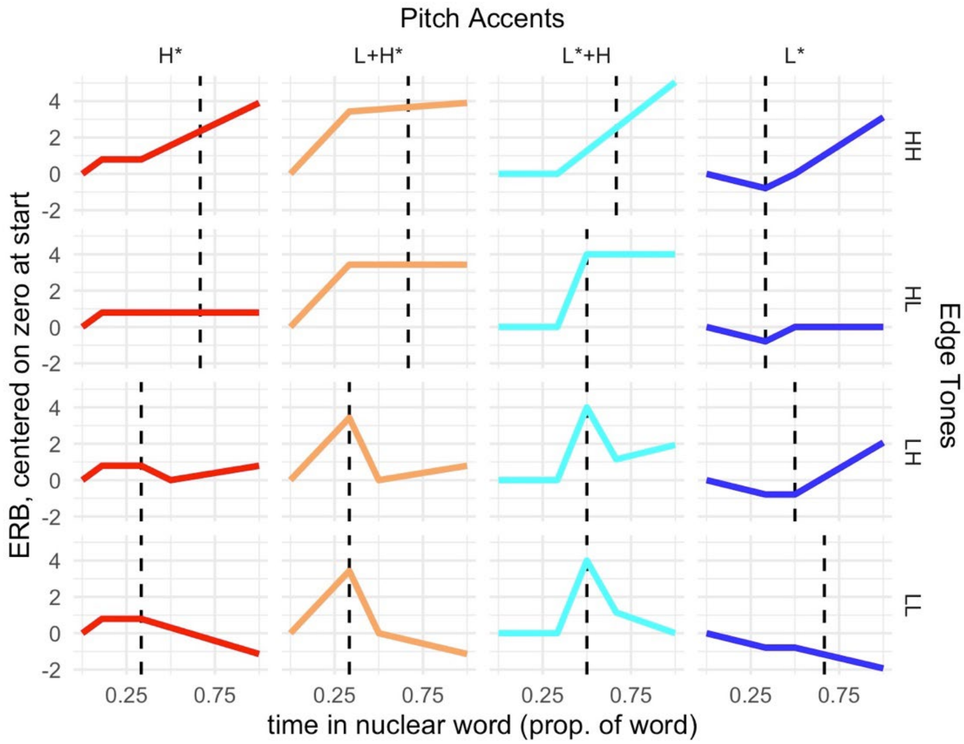

Figure 1: model trajectories for each pitch accent (columns) and each boundary tone (rows) in the corpus. The black vertical dotted line marks the end of the portion of the produced (imitated) trajectories that was used for modeling accentual F0 in this paper (see text for details).

3.2 F0 measurements and empirical F0 trajectories

In the study of speech dynamics, the value of F0 and its velocity are both important. Besides the value of F0 reached, the peak velocity (PV) and the time at which peak velocity is reached (TTPV) are also of paramount importance in the attempt to understand the dynamical model that generates the data (Perrier et al., 1988; Sorenson and Gafos, 2016; Iskarous and Pouplier, 2022. Though TTPV is not frequently used as a measure in intonation modeling, it serves to quantify the amount of delay in the location of a pitch extremum with respect to some supralaryngeal event like a stressed vowel, i.e., a measure of offset in the alignment of a tonal target. Indeed, the notion of alignment offset is one of the earliest innovations of the AM approach to tone, introduced by Goldsmith (1976) through the use of the star diacritic * to capture the offset in alignment of a tone to a tone-bearing unit (Goldsmith, 1974); we use TTPV to measure this offset.

Figure 2 provides a qualitative overview of the empirical pitch accent trajectories in our dataset. It shows the mean F0 trajectories (upper panels) and F0 velocity (lower panel) in the interval of the pitch accent (the initial 2/3 portion of the entire tune), for data aggregated across all subjects, and all tokens, including all three target words (i.e., names). The mean F0 trajectories are grouped by pitch accent (color), and the edge tone context (phrase accent followed by boundary tone; in panels). Here and throughout the figures in this paper, edge tone contexts are labeled with their two tonal elements, suppressing the diacritic features commonly used in AM annotation, e.g., the H-H% edge tone context is simply labeled HH.

Our first step in identifying a dynamical system model of our data is to note the patterns of variation in the empirical F0 trajectories. We observe variation across the pitch accents within each edge tone context, as well as variation for each pitch accent across the four edge tone contexts. Here we describe four measures of variation across F0 trajectories that will be relevant for our analysis: *F0 level at equilibrium* (maximum or minimum), *velocity of F0 rises* (i.e., slope) and *peak velocity* (the velocity maximum), *latency of peak F0 velocity,* and *F0 span* (Table 2). We examine these measures to establish and quantify the properties of F0 trajectories that differentiate the four pitch accents, H*, L+H*, L*+H, L*, which will inform the dynamical systems model.

Table 2. Empirical measures of F0 trajectories

Empirical measure	*description*
Level	The extreme (maximum or minimum) value along the F0 trajectory. The level is defined as the F0 maximum for rising trajectories, and F0 minimum for falling trajectories
Velocity	Amount and direction of F0 change across successive time steps, in plotted function of F0 over time, velocity corresponds to the slope of the F0 curve over an interval of time
Peak velocity	The value of maximum velocity along the F0 trajectory
Latency	Difference between the time at the beginning of accentual rise in F0 to time of peak velocity
Span	Difference in F0 extrema across the accentual interval: maximum - minimum for the rising pitch accents (a positive value), and minimum - maximum for L* (a negative value).

An immediately obvious distinction among the F0 trajectories of different accents is that they move toward different F0 maxima (rising accents) or minima (falling accents). These F0 extrema correspond to what are described as “F0 targets” in the intonation literature, and to equilibria of the dynamical systems under different values of the free parameter k, described in Section 4. We use the term **level** to refer to the extrema, maximum (rises) or minimum (falls), in an accent’s F0 trajectory. [3]

For accents with a rising F0 trajectory, we observe variation in the timing and extent of the rise. Variation in the timing of an F0 rise has long been known to be a crucial factor differentiating tonal categories, e.g., pitch movements with “early” vs. “late” peaks (Goldsmith, 1974; Bruce, 1977; Pierrehumbert, 1980; Pierrehumbert and Steele, 1989), and is evident in our data in the upper panel of Figure 2. In a dynamical systems analysis, the **velocity** function, which

[3] Here, we use the term “level” to refer to a particular F0 value (the F0 maximum or minimum) on a continuous scale of F0, within some interval. This is related to but not the same as the notion of “level” in opposition to “configuration”, as an analytic framework. In that usage, which pre-dates AM, “level” is closer to a phonological construct, with a discrete numeric representation, e.g., from lowest to highest, levels 1-4 (Trager & Smith, 1951), or a symbolic representation, e.g., the H and L tones of Autosegmental theory (Goldsmith, 1976; Pierrehumbert, 1980). See discussion in Ladd (2008a: 62-65).

characterizes the amount and direction of F0 change across successive time steps, provides a measurement of timing through the location of the **peak velocity**, as shown in the lower panel of Figure 2.

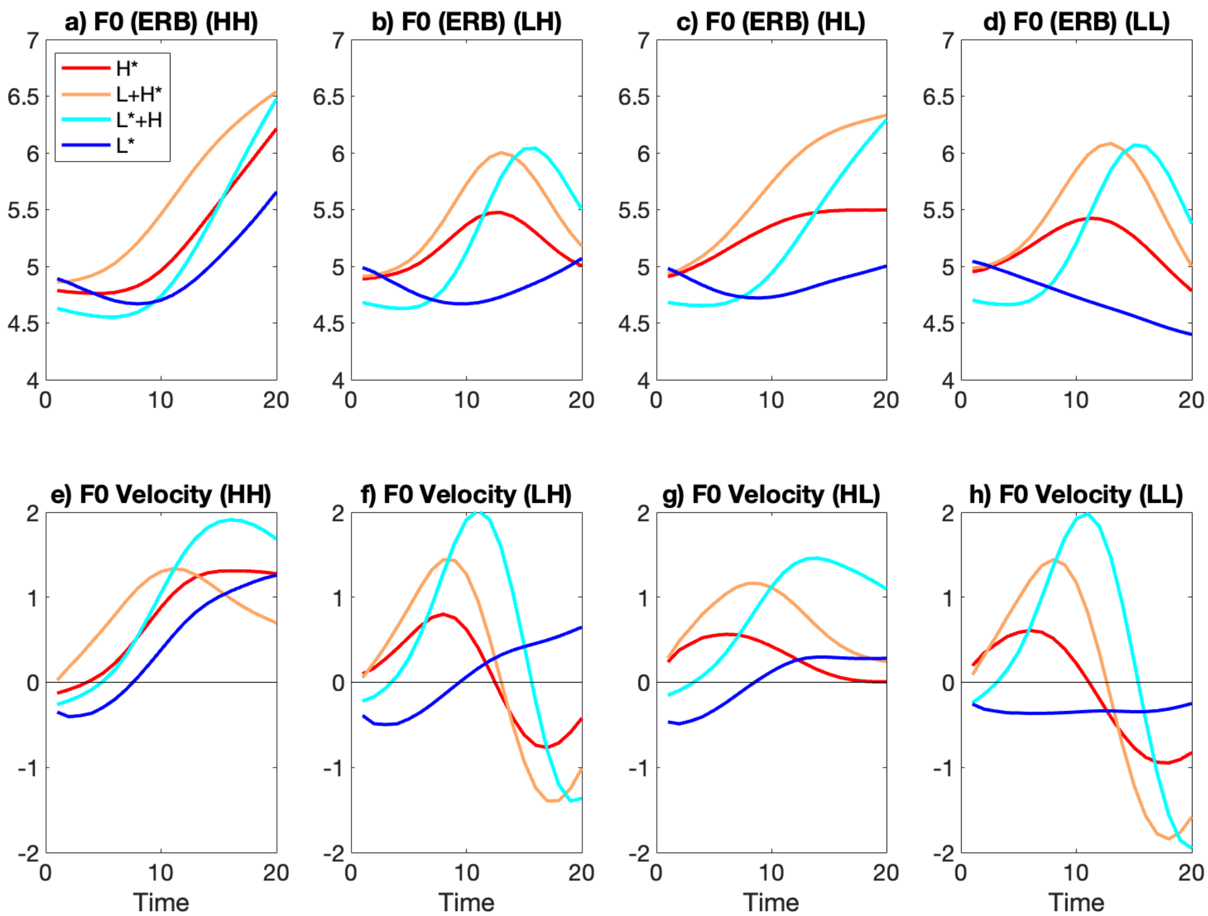

Figure 2. Mean F0 trajectories (upper panel) and mean F0 velocities (lower panel) for 4 pitch accents (H, L+H*, L*+H, L*) of imitated tunes in 4 edge tone contexts (HH, LH, HL, LL). The portions shown are 2/3 of the nuclear tune, showing the pitch accent (the topic of this paper), and a portion of the edge tones.*

The upper panels of Figure 3 show the distribution of **latency** values (TTPV) for each pitch accent, paneled by edge tone context. In examining latency variation across pitch accents, we focus first on the H-L% edge tone context, as this context exerts the least coarticulatory influence on the preceding accent, as this edge tone sequence specifies an F0 trajectory that simply maintains the final F0 value of the pitch accent interval up to the end of the intonational phrase (as described by Gussenhoven, 2005: 299-300 and Ladd, 1983: 721-759 2008; see also Ladd, 2008 128-129). For this reason, the question as to differentiation in latency among pitch accents is best answered by first examining the panel in Figure 3c. Latency increases across the rising accents: H* < L+H* < L*+H. For the bitonal accents (L+H*, L*+H), an ordering which corresponds to the difference in the temporal alignment of their F0 peaks as reported in prior work (Pierrehumbert, 1980; Beckman & Ayers, 1997; Arvaniti & Garding, 2007; Veilleux et al. 2006).[4] The same ordering of latency values across rising accents is seen in the L-H% and L-L% contexts, and in the H-H% context with the exception that the distribution of H* latency more closely resembles that of L*+H. This unexpected pattern for H* in the H-H% context can be understood in relation to the mean F0 trajectory for H* in Figure 2a, which is very similar to L*+H, i.e., speakers appear to have neutralized the contrast between H* and L*+H in just this edge tone context (imitations of the H* stimulus produced with F0 trajectories that more closely resemble the L*+H stimulus; see Steffman, Shattuck-Hufnagel & Cole, 2022 for discussion). We assess the magnitude of the differences in F0 latency among the rising pitch accents using Cohen's d (Cohen, 1986), a measure of effect size (here, the effect of pitch accent category on F0 measures). The upper panels of Figure 4 shows Cohen's d for Latency for each pitch accent pair, paneled by edge tone condition.[5] Cohen (1986) provided the heuristic that (taking the absolute value of d), d ~ .2 is a small effect, d ~ .5 is a medium effect, and d ~ .8 or above is a large effect. Pairwise differences that are medium or above based on absolute d values are displayed in bold red. We are interested in how latency differentiates the three rising accents (with only one low/falling accent in our dataset, there is no opportunity to observe relevant differences in the latency of the F0 minimum for L*). There are large differences between the bitonal rising accents L+H* and L*+H in all edge tone contexts, and similarly large differences between L*+H and H* except in the H-H% context, as already noted. Setting aside the H-H% context, the pair H* and L+H* show a medium difference only in the L-L% edge tone context. Overall, and again setting aside H* in the H-H% context, we observe substantial differences in latency, as measured by TTPV, of rise timing for the rising accents in our data.

The lower panels of Figure 3 show the distribution of **F0 span** for each pitch accent, paneled by edge tone context, and Cohen's d for accent pairs in Figure 4 (lower panels). L*+H and L+H* have a larger F0 span than H* before L-H%, H-L%, and L-L%, with medium-to-large effect size, while L*+H has a larger span than L+H*, though the effect size is small. These results are consistent with the descriptions of these accents in the literature (e.g., Arvaniti & Garding, 2007; Beckman & Ayers, 1997; Burdin et al., 2022; Veilleux et al., 2006), where the H* rise begins earlier and is more gradual than the rise for the bitonal accents, while for the bitonal accents the rise starts from a low pitch at the onset of the stressed syllable (L+H*) or later (L*+H). We note that the characterization of rising accents in MAE in terms of differences in their F0 trajectories is the subject of much debate in the literature (Calhoun, 2012; Ladd & Schepman, 2003; Ladd, 2022; see also discussion in Ladd 2006: 96-97, 136-137). Most relevant for the discussion here is that experimental findings from prior work show little difference in F0 at the rise onset for H* and L+H* (Arvaniti & Garding, 2007; Calhoun, 2012; Dilley, 2005; Ladd & Schepman, 2003). In fact,

---

[4] These works do not describe a distinction in peak alignment between H* and L+H*, for which the primary difference is claimed to be the presence of a low pitch target at the onset of the rise only for L+H*, but visual inspection of examples in the ToBI tutorial (Veilleux et al., 2006, section 2.5.2.2) suggest that an earlier F0 peak for H* compared to L+H* occurs in at least some phonological contexts.

[5] Cohen's d is a measure of effect size that can be used to quantify the difference between distributions. It measures the difference between the means of two variables normalized by their summed standard deviation: $\frac{\mu(x)-\mu(y)}{\sqrt{\sigma(x)^2+\sigma(y)^2}}$. We calculate Cohen's d with x and y as the F0 measures of latency (Time to PV) and span for two pitch accents, x and y.

our evidence is similar. While the mean trajectories in Figure 2 show a substantial difference between the F0 maximum for L+H* and H*, there is at best a very small difference in the initial F0 values for these two accents. Differences in F0 span in our data derive from the difference in F0 maximum of H* and L+H*, and the lower initial F0 of L*+H. Together, these differences yield the observed ranking of accents by F0 span: H* < L+H* < L*+H. Figure 5 shows the distributions of initial and final F0 in each pitch accent, after centering each trajectory around its mean to partially normalize for overall F0 differences across participants, and Figure 6 shows the accompanying Cohen's d's. We see that the distributions of the initial F0 value are quite similar across all four edge tone contexts, but that in L-H%, H-L%, and L-L% edge tone contexts, there is an ordering in the final F0 value: H* < L+H* < L*+H. Differences in the final F0 value for rising accents in the H-H% are much smaller, presumably reflecting a strong coarticulatory influence of the very high F0 at the end of this tune (not shown; for details see Cole et al. 2023; Steffman et al., 2022).

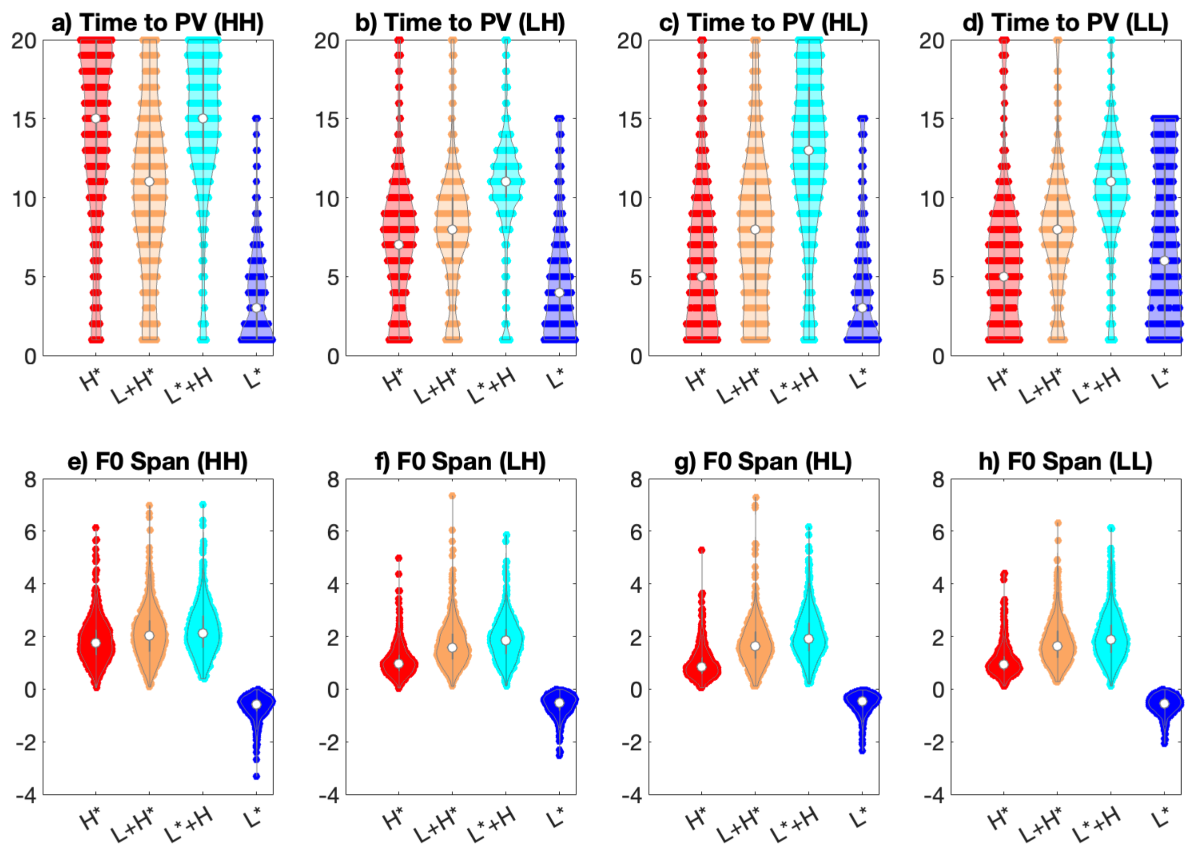

Figure 3. Time to peak velocity and F0 span by pitch accent and edge tone context.

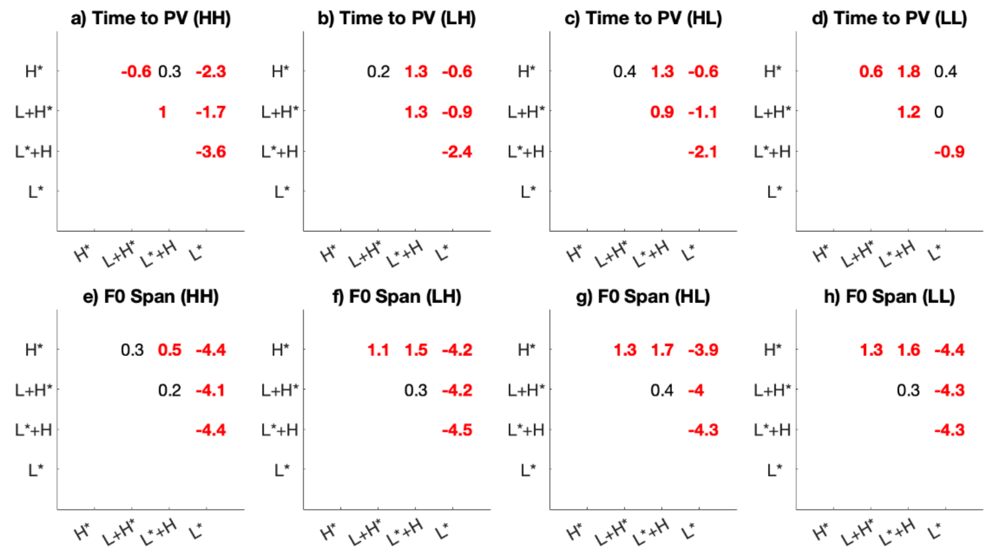

Figure 4. Cohen's d for distinctions between pitch accents in F0 latency (time to peak velocity) and span (see Figure 3). Number in red indicate medium or large effect sizes.

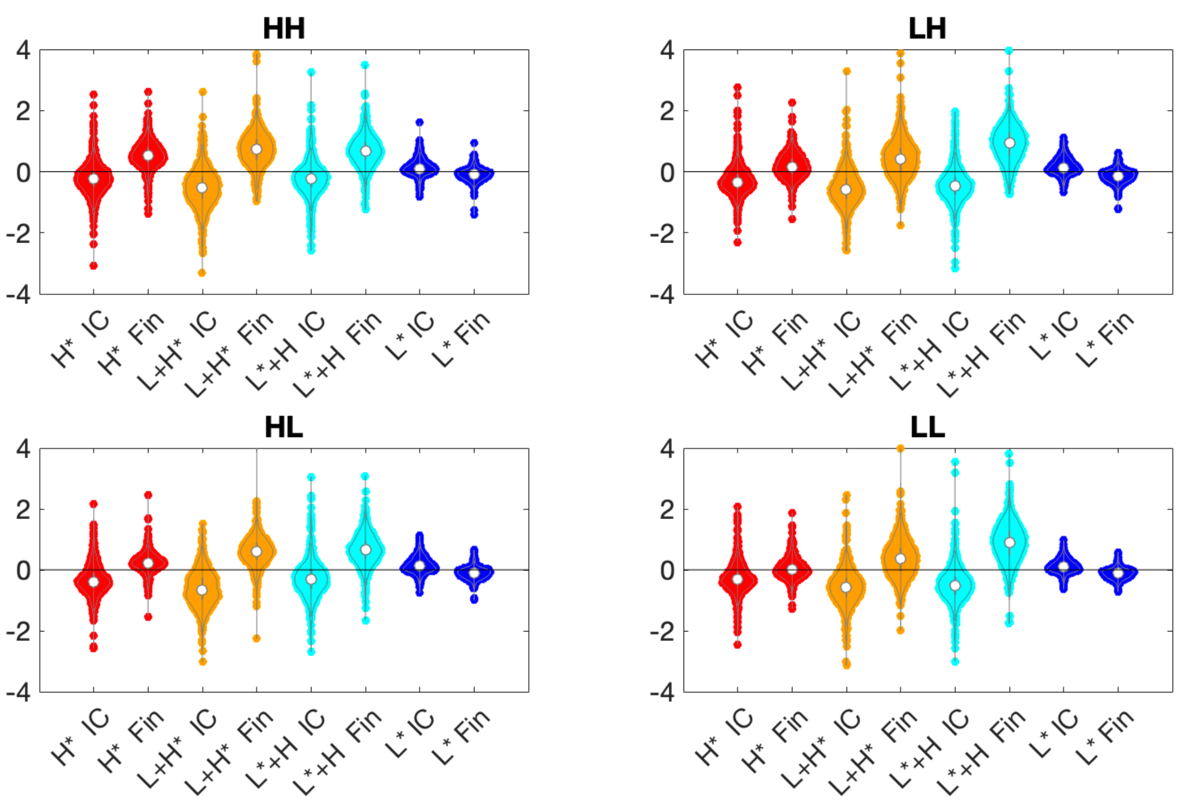

Figure 5. Initial and Final values of centralized F0 trajectories by pitch accent, and each of the edge tones.

To summarize the results so far, among the rising accents there is a gradation across accents in both the latency (Figure 3, top) and magnitude (Figure 5) of the rise, which yields an ordering of the accents: H* < L+H* < L*+H. This ordering can be described as *rise later, rise higher*. Comparing latency and velocity (Figure 2, bottom) yields the same ordering, which can be described as *rise later, rise faster*. In Figure 6 we assess these relationships directly, comparing latency with F0 span, F0 level (= max F0) and peak velocity from left-to-right in the first three columns. Here latency values are binned in quarter-portions of the temporal accent interval on the x-axis, from 1st to 4th quarters. We also assess how peak velocity relates to F0 span and F0 level in scatterplots shown in the fourth and fifth column. The scatterplots show all rising pitch accents pooled together, so the results in Figure 6 are true of individual trajectories, regardless of the specific pitch accent label that represents the target of the imitated production. Despite some variation in the relation between latency and the other three variables, there is an overall positive trend: As latency increases across the four temporal quarter-intervals, F0 span, F0 level and peak velocity trend upwards.

Each panel of columns 1-3 in Figure 6 shows Cohen's d for the distributions of F0 span, F0 level, and F0 peak velocity at the 1st through 4th quarter interval, as a partial quantification of the size of the latency effect on those measures, i.e., quantifying the strength of the positive relationship between latency and each of those measures. All these Cohen's d values are large, indicating that this is indeed a robust pattern: the later the rise, the higher and faster the rise. The scatterplots in Figure 6, columns 4-5, show Pearson's correlation coefficient, which similarly quantifies a strong relationship between peak velocity and F0 span/level. These data speak directly to one of the main debates in intonation research, the *level-vs.-configuration* debate. The dynamical approach relates levels of variables to their rates of change, so this type of debate does not make much sense in the context of dynamical theory. Indeed, a dynamical theory analysis predicts a relationship between the level reached in a pitch accent and the maximal slope of rise (i.e., peak velocity). The scatterplots in Figure 6 show this quite robustly for the rising accents: there is a strong relation between the peak velocity of rise and the change in F0. The appreciable Pearson correlation coefficients show, fundamentally, that F0 measures of level and configuration are not orthogonal.

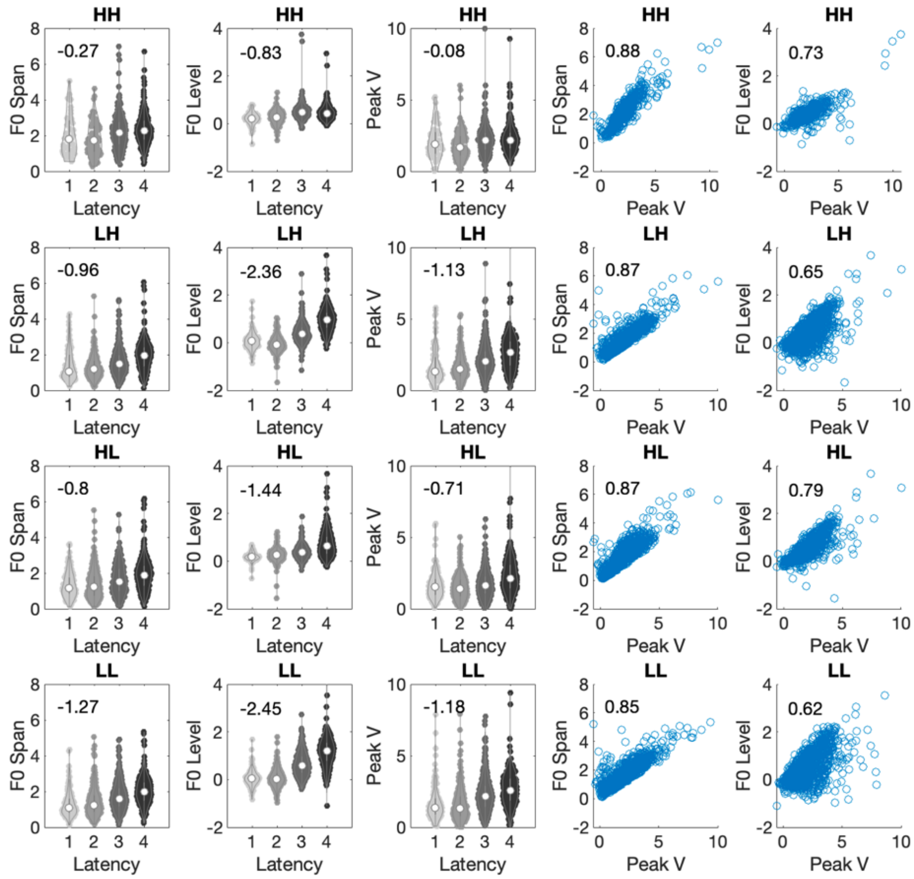

Figure 6. Relationships between paired F0 measures for rising pitch accents: latency (Time to peak velocity), F0 span, F0 level and peak velocity. Column 1) rise latency and F0 span; Column 2) rise latency and F0 level; Column 3) rise latency and peak velocity; Column 4) peak velocity and F0 span; Column 5) peak velocity and F0 level. Data aggregated over three rising accents: H, L+H*, L*+H. Cohen's d for F0 span, level and peak velocity distributions is shown in the upper left of each panel in columns 1-3; Pearson's r in the upper left of each scatterplot.*

One more aspect of pitch accents that we would like to highlight is the average initial velocity of F0 in the pitch accents, measured over the first 20% of the pitch accent, since this provides information about the distinction between the monotonal H* accent and the L initial tone in the bitonal rising accents . Figure 7 provides kernel density estimates[6] of the distributions of the velocities. As expected, most of the values of initial velocity for L* are negative, but there are also quite a few negative velocity initiations for the rises, with the most for L*+H and least for H*. We believe this reflects the "scooped" shape seen most often in productions of the L*+H

[6] A kernel density estimate is the estimate of a probability distribution that underlies an empirical distribution (here, F0 measures).

accent, described in work pre-dating AM (Vanderslice & Ladefoged, 1972: 822) and in the AM framework (Ladd, 1980; Gussenhoven, 1984; see also discussion in Ladd, 2008a: 94). The initial negative velocity for some rises is an extreme form of scooping, where there is a fall before the rise. As we show below, this aspect of the data is important in determining the underlying dynamic of pitch accents, since the mathematical theory should be able to predict rises that only rise, as well as rises that begin with a fall before the rise onset.

In the next section, we motivate a dynamical system that generates the pitch accents of MAE, capturing the dynamic F0 patterns related to the timing (*velocity, latency*) and extent (*span*) of the F0 movements that differentiate accents in our data, as described above. Although we use the results above as a quantitative springboard into the model, many of the observations we try to account for in the model have been made previously in the literature on MAE intonation, including works cited above. Before launching into the dynamical model, we first summarize the key elements of the AM theory used in the analysis of MAE pitch accents for modeling the proposed phonological differences among accents (Table 3), focusing also on how these constructs fare in capturing the empirical measures presented in Table 2 above.

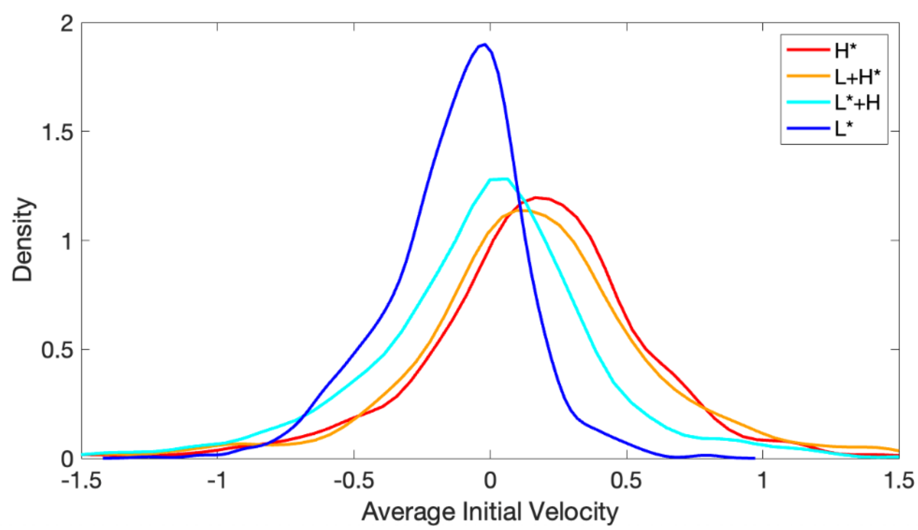

Figure 7. Distributions of the average initial velocity of F0's for each pitch accent pooled over all edge contexts.

Table 3. AM constructs for the analysis of accentual F0 movements

Construct	*Example with reference to AM*
Tone level	High, Low
Linearity	Tone sequences in bitonal accents, e.g., L+H*, L*+H, H+!H*
Alignment	Early peak (L+H*) vs. late peak (L*+H)

In the AM model of MAE pitch accents (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986), three constructs are involved in defining phonological contrasts among accents: tone level, linearity, and alignment. The construct of tone ***level*** was already present in earlier approaches to the analysis of English intonation (Pike 1945; Trager & Smith, 1951; see discussion in Ladd 2008a: 62-75), though AM restricts the number of levels to two, Low (L) and High (H), with downstepped High (!H) added in subsequent work (Ladd 1983). These level tones are the primitive features that combine to define dynamic patterns, or configurations, of rising and falling pitch. Configurations arise from tone concatenation due to ***linearity,*** a construct that imposes a linear precedence relation between the tones within a prosodic domain. Tones combine to form linear sequences across phonological anchors such as syllables, moras or segments, but in bitonal accents, tones may also combine in a linear sequence on a single, shared anchor, for which AM uses the representational device of the plus symbol (+), e.g., L+H. The ***alignment*** construct is critical for the analysis of bitonal pitch accents, to designate one of the two tones as aligned to a phonological landmark—the stressed syllable. The aligned tone is marked with the star diacritic (*), and the other tone in the bitonal accent is sequenced to precede (L+H*, H+!H*) or follow (L*+H) the starred tone but is not phonologically aligned to a segmental or syllabic landmark.[7]

Combinations of these three constructs are sufficient to encode a basic, level distinction between accents with high vs. low pitch targets and some of the most salient timing differences among accents. For instance, timing differences in the realization of High-tone targets (F0 maxima) for rising accents are captured through the combined effects of linearity and alignment. *Linearity* allows the composition of tones in bitonal accents, which differentiates dynamic rising (L+H*, L*+H) and falling (H+!H*) accents, anchored at the beginning or end of their F0 movement with a stressed syllable, from the monotonal accents (L*, H*), which specify only a single F0 target, and thus do not restrict the F0 movement into or out of that target. *Alignment* distinguishes the bitonal rising accent with an early peak, L+H*, from the late peak accent, L*+H, and also specifies that the sole falling accent, H+!H* is realized with a falling F0 movement that reaches its target, a downstepped high tone, in the stressed syllable.

What AM offers, fundamentally, is a model based on linearly sequenced level tones, some of which are aligned to phonological landmarks, phonetically implemented through the context-sensitive specification of F0 values as tonal targets with linear interpolation between successive targets. But, as we have mentioned earlier, this model doesn't capture all the salient properties of accentual F0 movements. For instance, in rejecting earlier approaches to English intonation that center configurations as the primitive units of analysis, (Bolinger 1951; O'Connor & Arnold, 1973; t'Hart & Cohen, 1973; see discussion in Ladd, 2008a; Arvaniti, 2011), AM fails to capture certain observed distinctions in the shape of F0 movements, e.g., the extent of an F0 maximum (peak vs. plateau), or in rise curvature (D'Imperio, 2000; Barnes et al., 2012).[8] With respect to our MAE pitch accent data, the AM model offers no explicit encoding of the pattern of scooped

---

[7] The notion of tonal alignment to a phonologically designated anchor predates AM. For example, alignment (under the term "timing") distinctions that underlie the Swedish accent system are discussed by Bruce (1983), citing work dating back to Haugen (1949). The use of the star diacritic as a representational device marking alignment was introduced by Goldsmith (1976) in his proposal for Autosegmental Phonology and adopted in Pierrehumbert (1980) for the specification of bitonal pitch accents in MAE.

[8] Pierrehumbert (1981) argues for an F0 implementation model with "sagging" interpolations between successive High tonal targets, and in the fall from a High to a Low target. But for the bitonal rising accents, the transition between and initial Low tone and the following High tone target is proposed to be monotonic, never dipping below the lower value (p. 990). This model does not generate a scooped rise, or a distinction between scooped and domed rises (Barnes, et al. 2012).

F0 rise in productions of the L*+H accent, a property related to rise configuration. It also does not predict or model the observed variation in the scaling of tonal targets (F0 maxima) among the rising accents (Figure 6),[9] or the relationship in rising accents we described earlier as "*later-higher-faster",* i.e., the relationship between the latency and scaling of F0 maxima, and F0 velocity.

Before leaving this overview, we remark on one final property of accentual F0 movements that is pervasive in our data: There is significant *F0 variation* in the production of pitch accents in MAE, including within-category and within-speaker variation, which is evident in the medium effect sizes reported above (Figure 4). We observe a similar degree of variability in pitch accent production within and across discourse contexts, speech styles, and speakers, in our prior experimental and corpus work on MAE (Chodroff & Cole, 2018, 2019; Im, Cole & Baumann, 2023), and similar variation is reported for German (Grice et al., 2017; Röhr, Baumann & Grice, 2022).

In Section 5, we will discuss other properties of pitch accent systems, and tone systems more generally, that modifications of the proposed dynamical model could potentially account for.

4. Combinatorial Dynamics for a Pitch accent Model

The goal of this section is to introduce a dynamical framework for describing pitch accents of MAE in particular, but we also seek to lay the foundation for describing pitch accent systems in general. Our exposition will be combinatorial in nature, first describing dynamical atoms—simple dynamical behaviors due to one-term differential equations (a relation between a value and its change), then we show how combining these atoms yields differential equations for pitch accents. This approach also allows us to show how the empirical measures observed in our data, related to *level, velocity, latency, span* (Table 2), evolve from their simplest precursors. Our argument that the model we will provide is a *minimal* one is based on our combinatorial/hierarchical approach, since we start with the simplest dynamical system atoms, building just enough structure, gradually, till we get to minimal system that can account for the dynamical properties. This approach also allows the exposition to be self-contained, but the reader can also consult Sorenson and Gafos (2016), Roessig et al., (2019), Iskarous (2017), and Iskarous and Pouplier (2022) for other accessible introductions to dynamical systems analysis. The main dependent variable for pitch accents, F0, will be labeled F.

[9] More generally, AM does not offer a model of systematic variation in the scaling of tonal targets within the local tonal span, or register. See discussion in Ladd (2008a: 188-210).

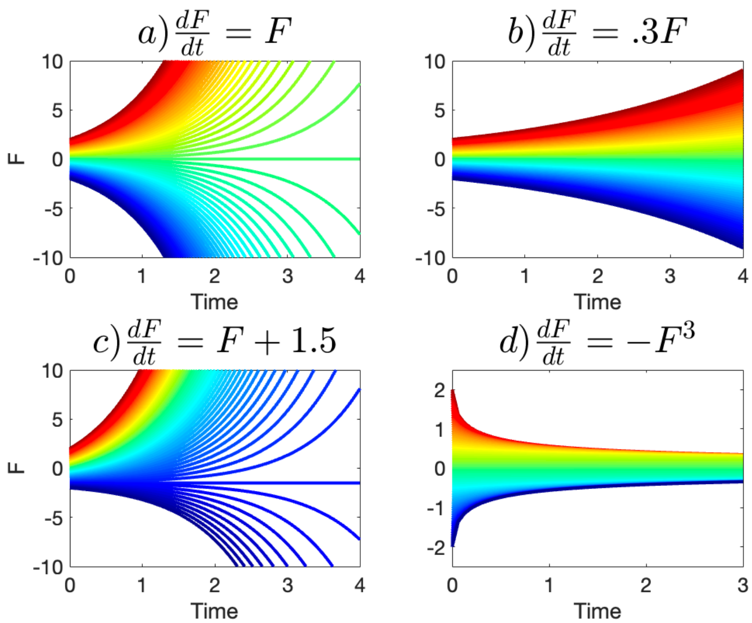

Figure 8. Dynamical atoms: a) positive feedback system with k = 1, boundary = 0, unstable rise configurations above and unstable fall configurations below; b) positive feedback system with k = .3, boundary = 0; c) positive feedback with a boundary = -1.5; d) cubic negative feedback system. Colored trajectories correspond to initial values of F (the IC) between -2 and 2.

4.1 Dynamical Atoms

Differential equations are about change, and we begin by describing four motifs, or atoms, of change, which will be later combined. The left-hand side of a differential equation refers to the change in a dependent variable's state or value, here F, with respect to some independent variable, here time t. The (potentially infinitesimal) change in F with respect to t will be symbolized by $\frac{dF}{dt}$. If F is plotted with respect to t, $\frac{dF}{dt}$ is the slope of F at time t. The right-hand side of a differential equation defines the value of $\frac{dF}{dt}$ through some function, usually of F itself. What we seek are functions F(t), F as a function of time, that make the differential equation true. For instance, if a differential equation says that the value of the derivative (graphical slope) at every point should be equal to the value of the function minus the cube of the value of the function, then the functions that we seek are the only ones with that property: that their slope at every point is equal to the value at the point minus the cube of that value. These functions are the *solutions* of the differential equation. The simplest differential equation where the change in F, $\frac{dF}{dt}$, depends on its current value positively is $\frac{dF}{dt} = F$, which represents a positive-feedback loop: if the value of F starts out as a positive number, the differential equations says that the slope should be positive, so F increases further increasing the slope, etc. This is called a positive feedback system since the value positively feeds the slope, which feeds the value, etc. Such solutions can be seen in Figure 8a in the top half of the graph. Each curve is for a different initial positive value of F. If F starts as a negative value, then F falls, and keeps on falling, as seen in the bottom half of Figure 8a. These are examples of the *exponential function* solutions with the initial condition (IC), the value of F at t = 0, varying from 2 (red) to -2 (blue) . The curves go to +∞ for IC's above 0, to -∞ for IC's below 0, and stay at 0 if the IC is exactly 0. Through this dynamic, $\frac{dF}{dt} = F$, 0 becomes a very special value of F, an *equilibrium* value, since if F starts exactly at 0, the system is in equilibrium and remains there. The value 0 is an *unstable* equilibrium, since the slightest deviation from 0 leads to explosive deviation from equilibrium. For $\frac{dF}{dt} = F$, dynamicists call 0 a repellor, since it repels the state F. In the context of phonetics and phonology, 0 can be thought of as a *boundary*, as it divides all the possible values of the IC and the entire set of F-curves into two classes, the rises above 0 and the falls below 0, through the dynamic of repulsion. It may seem strange that a discrete notion like *boundary* can be discussed in the context of dynamical states like F that can take on any possible continuous value, by using continuous mathematics. This is possible because differential equations can *discretize* continuous state spaces into regions, with boundaries or thresholds (Strogatz, 1994; Abraham and Shaw, 1992). Therefore, we already see something useful arising from this simple equation: the notion of boundary. Another property arising from this equation is the distinction between rising and falling F0 configurations. Rises and falls result from the repulsion from an unstable equilibrium, 0 in this dynamical system. The rises and falls in Figure 8a are not the same as those observed in actual F0 contours, as this model is too simple; however, we are trying to show the source of each property in simple models before moving to more complex models where those properties persist, and new ones emerge. This model already demonstrates the *possibility* of rise and fall configurations. As the right-hand side becomes more complex, we will see that realistic rise and fall configurations will be possible, as well.

We turn next to consider latency, another of the empirical measures in Table 2, which in our data distinguishes among rising pitch accents that differ in the timing of the F0 maximum. This property relates to the AM construct of alignment (Table 3). The simplest differential equation $\frac{dF}{dt} = F$, where a variable positively influences its own slope, already has implications for the timing of rises and falls. Consider the timing of rising/falling trajectories as a function of the distance between an IC and the equilibrium in Figure 8a. When the IC is large positive or negative (red and blue curves), leading to a large distance between IC and the equilibrium at 0, the initial slope is large in magnitude and therefore the function rises/falls fast and early. In contrast, when the IC is closer to the equilibrium, it takes a long time for the slope to rise enough for divergence from equilibrium to take place (green curves), a behavior we describe as *stickiness*. The closer an IC is to the equilibrium, therefore, the longer (in time) F sticks to the equilibrium, and consequently, the greater the delay in its rise/fall. Therefore, this very simple system also reveals a source for the possibility of delay in the onset of a rise or fall, which we argue is a dynamical root for variation in latency (equivalently, variation in alignment). This is because, for variation in latency to occur, F0 trajectories must have temporal malleability leading, e.g., to an earlier vs. later rise or fall as attested in some of the seminal work in AM theory (Goldsmith, 1974; Bruce, 1977). The specific association we have from this simplest dynamical system is that the smaller the distance between IC and the (unstable) equilibrium, the later the rise/fall. The specifics of latency in empirical F0 trajectories are not modeled correctly by this differential equation, but the distance between IC and equilibria will still play a role in more complex differential equations that will be argued to more accurately model the empirical F0 trajectories.

If the simplest differential equation just discussed is modified to have a variable positive coefficient *k*: $\frac{dF}{dt} = kF$, then k changes the repulsive power of the boundary. If k is large, then F influences its own change by a higher factor, so solution curves are repelled from 0 faster and greater. The closer k is to 0, on the other hand, the slower and less the rise/fall. This can be seen for k = .3 in Figure 8b. Therefore, the value of k is a second dynamical determiner of stickiness to the boundary, with low k leading to later latency of rise/fall and high k to earlier latency of rise/fall.

A different modification of $\frac{dF}{dt} = F$ is to add a constant C to it: $\frac{dF}{dt} = F + C$. Figure 8c shows solutions when C = 1.5. F still influences itself positively, so we still get repulsion from the boundary. The difference here is that the boundary is no longer at 0, but at -C, so for this example, the rise/fall boundary is at C = -1.5. This is obvious, since if the IC = -1.5, then $\frac{dF}{dt} = -1.5 + 1.5 = 0$, and the F value making the right-hand side 0 is the equilibrium. In general, for this system, the boundary, which we will call F^{equi}, is set to equal -C. Therefore, C variation changes the distance between the IC and the boundary, since for a fixed IC, C changes the boundary location. Earlier we saw how the distance between the IC and equilibrium leads to stickiness, and how that can be accomplished by varying the IC. Here we see that the distance can be varied by changing C, a third factor that can alter the latency/alignment of rises and falls.

Another simple form of self-dependence (i.e., the change in F depends on the current value of F) is negative feedback, where a positive value of F leads to a negative change, and a negative value of F leads to a positive change. In such systems, if an event pulls F away from the equilibrium, F subsequently trends back towards that equilibrium. One differential equation that leads to this behavior is $\frac{dF}{dt} = -F^3$. A positive F will lead to decay ($\frac{dF}{dt} < 0$) and a negative F will lead to growth ($\frac{dF}{dt} > 0$), as can be seen in the solutions in Figure 8d. Again here, an IC of 0 leads to no change, so 0 is an equilibrium of this equation. However, small deviations from 0 will be reduced, not amplified as we saw for $\frac{dF}{dt} = F$. Therefore, 0 is another discrete entity, called a *stable* equilibrium or an attractor, picked out as a special state amongst the infinite set of possible real F values. Fowler et al. (1980) and Browman and Goldstein (1989) associated attractors with contrastive values in phonology, since discrete stable values are *category*-like, where attractiveness is the basis for the regularity of the linguistic behavior. The initial value for such a dynamical system is the state before speaking or at the end of some earlier contrastive unit, and the trajectory towards the attractor is the spoken actuation of the contrast.

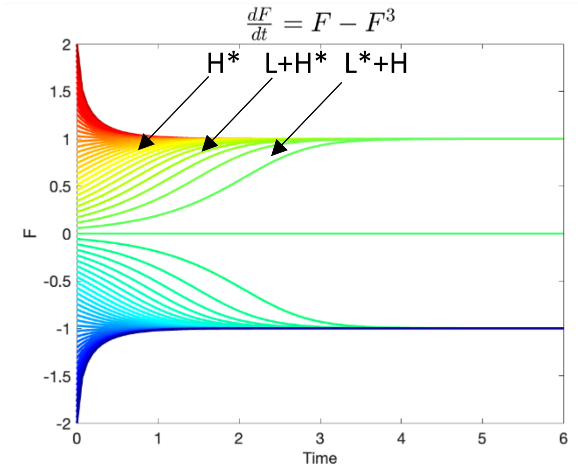

Figure 9. Dynamic Molecule 1: Linear positive feedback and cubic negative feedback, leading to 2 stable equilibria/levels, with an unstable equilibrium (= boundary) in between. Initial conditions are variable. The pitch accent labels are explained in the text.

4.2 Dynamical Molecules
4.2.1 Equilibria and latency

Now consider what happens when we combine the two dynamical systems capable of repulsion and attraction into one: $\frac{dF}{dt} = F - F^3$, whose solutions are shown in Figure 9. If F is initially a small positive number less than 1, then F^3 can be ignored, since it's much smaller than F. So we end up with $\frac{dF}{dt} = F$, and F starts to rise exponentially—however, as F gets larger, F^3 can no longer be ignored. And when F reaches 1, $\frac{dF}{dt} = F - F^3 = 1 - 1 = 0$. So 1 is a stable equilibrium, while 0 is still an unstable equilibrium. If F starts as a small negative number between 0 and -1, it will be attracted to F = -1, another attractor. Therefore, this system has three equilibria, three special discrete values along the F-continuum that impose a great deal of structure on it. In the context of phonetics and phonology, we can regard this system as having 2 stable *categories,* a high valued category (1), and a low valued category (-1), and a boundary (0) in between. As opposed to the rise and fall configurations in the simplest system we considered, the rise and fall configurations of this system level off. We posit that this is a dynamical precursor for the notion of *level* in the AM theory, giving rise to H(igh) and L(ow) as stable equilibria of a dynamical system.

The question now is how $\frac{dF}{dt} = F - F^3$ can be complexified to predict the latency differences between pitch accents, as well as other properties that accompany latency. We have already seen that there are three factors that can lead to differences in rise latency: 1) the value of the IC; 2) the value of a boundary C; 3) the value of k. Since we are seeking a minimal model of pitch accent systems, we will first investigate models where there is variation in only one of these parameters (IC, C, k), while the other two are held constant, while also keeping the model with as few terms on the right-hand side as possible. The parameter that varies would then represent differences in pitch accent. There are therefore three basic hypotheses about which parameter corresponds to proposed phonological pitch accent distinctions.

We've already seen the first hypothesis: $\frac{dF}{dt} = F - F^3$, with varying initial conditions as in Figure 9, needing no further complexity. This dynamical system models differences among pitch accents by varying the IC. The closer the IC is to the boundary 0, the stickier the F0 trajectory, and the slower the trajectory arrives at the value of a stable equilibrium. One could imagine, then, that the L*+H accent, with high latency, could be modeled by an F-curve that has a positive IC close to 0; L+H* with a higher IC, and H* with an IC closer to 1. Indeed, there is a notable similarity between the labeled pitch accents in our Figure 9 and empirical F0 trajectories of the same rising accents in Pierrehumbert 1980.[10] Moreover, early rising H*, starting at a higher IC, shows a smaller span of F0, while, at the other extreme, L*+H, with late rising, shows a much larger F0 span. This would suggest that the very simple F0 model $\frac{dF}{dt} = F - F^3$, the same model from which levels emerge, also generates timing differences between the rises as well as F0 span differences. L* in this model would, of course, be based on an IC below 0. However, differences among the rising accents in the scaling of their F0 maximum value, which is observed in our data (see Figures 2a-d and 3), as well as in data of Arvaniti and Garding (2007) and Burdin et al. (2022), are not predicted by this model. In these data, H* has the lowest F0 maximum, while L*+H has the highest. We believe that even if there are systematic differences among pitch accents in their IC values, such differences may not be the most crucial; instead, variation in C or k may lead to a better model that captures both timing and scaling differences among the rising accents. Therefore, as we approach a more realistic model of F0 trajectories, we will take IC to be constant, the assumed *baseline* for a speaker, from which rises and falls can be made. Later we will discuss what happens when we vary IC in addition to C or k. In addition, we will (for now) assume that this baseline is a small negative value, signifying that before the pitch accent begins the system is inhibited.

[10] The reader is directed to these figures in Pierrehumbert 1980: From chapter 2, Figures 2.22 illustrates the later rise onset for L*+H vs. H*. From chapter 4, Figures 4.3 and 4.4 illustrate the later rise onset for L*+H compared to L+H*, while Figures 4.30 and 4.31 illustrate the later rise onset for L+H* compared to H*.

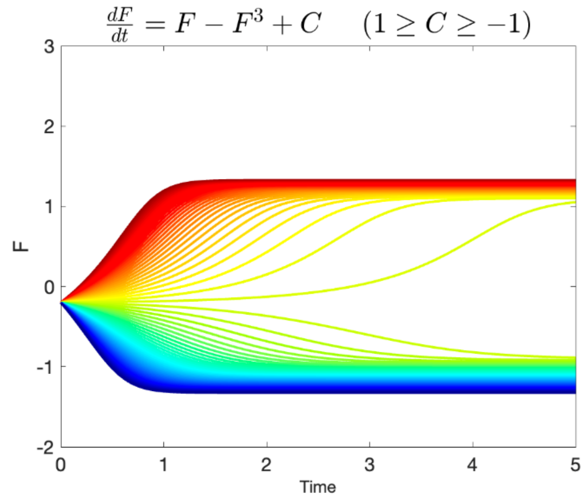

Figure 10. Dynamical Molecules 2: initial condition IC is fixed at F = -.2 and k is fixed at 1, while C varies.

We turn now to the second hypothesis, that pitch accent variation is modeled through variation in C. We enrich our model by adding a constant C to obtain $\frac{dF}{dt} = F - F^3 + C$, and test whether varying that constant yields a possible model of pitch accent variation. In the enriched model, IC is fixed at baseline and k, the coefficient of the linear term, is fixed at 1. Since the goal is to model different pitch accents, and C is the only variable here, (IC and k are fixed), the current hypothesis is that each value of C picks out a different pitch accent. We have already discussed how varying C varies the location of the boundary (compare Figures 9a and 9c), and since we are now assuming that the IC has a fixed value for all trajectories, the IC will be either above the boundary or below the boundary, depending on its fixed value. Figure 10 shows trajectories generated from this system, all emanating from IC = -.2, with one solution for each fixed value of C. As expected, the trajectories rise when the IC is above the boundary value and fall otherwise. Also, since the distance between the IC and boundary changes with C, we expect to see some trajectories rising earlier than others (red-early, yellow-late), which is confirmed in Figure 10, where we also observe trajectories that rise early rise to a higher F0 maximum, while those that rise late achieve a lower maximum. This is opposite to the pattern observed in our empirical data, where we observe that trajectories that rise early also have a higher F0 maximum, i.e., *rise-early-rise-higher*, discussed above in terms of the ordering of rising accents for latency (Figure 3, top) and magnitude (Figure 5). We therefore abandon this model.

4.2.2 Emergence of levels, alignment and linearity

Our third hypothesis is that pitch accent variation is due to k-variation, while IC and C are fixed. The differential equation that defines this dynamical system is given in (1). In the terminology of dynamic analysis, this kind of system is called an *imperfect pitchfork bifurcation* system (IPB; the bifurcation is illustrated in Figure 11).

$$\frac{dF}{dt} = kF - F^3 + .5 \qquad (1)$$

Figure 11a shows solutions of the IPB system in (1) for different values of k. Keeping the IC of F at a small negative number, when k is a small positive number, both kF and -F^3 will also initially be very small. In that case the dominant term is C = .5, therefore $\frac{dF}{dt} > 0$, which is the main condition for generating rising trajectories. In short, setting k to a small positive value results in rising trajectories. A different pattern emerges when k is set to a large positive value. Still keeping the IC at a small negative value, kF will initially be large and negative (positive k multiplied by negative F), exerting a negative force on the change in F which eventually (and soon) overcomes the rising force from the positive C. But the negative divergence from kF is then balanced by -F^3 (which will be large positive) to get falls that resemble L*. Within the rises, as k increases, the balance between the rising force of C = .5 competes with the falling force of kF-F^3. The kF- F^3 force for falling delays the rise more and more, as the force for falling increases. This can be seen in Figure 11a where an increase of k (from red to green) leads to a later and later rise. We therefore see how under this hypothesis there is a balance between the linear positive feedback (kF) and cubic negative feedback (-F^3), leading to configurations that level off, but there is also an interlinked balance between those two terms and the rising force of positive C that results in variation in timing. Moreover, this dynamical system, hypothesized to underlie the pitch accent system of MAE, predicts the empirical measures accompanying latency. H*-like trajectories that rise earlier and slower also rise to lower F0 maxima, whereas L*+H-like trajectories that start later and faster rise to higher peaks.

To understand the system in (1) at a deeper level, we will plot what dynamicists call a bifurcation diagram as in Figure 11b. For each value of k, we compute the value(s) of F that leads to $\frac{dF}{dt} = 0$. These are the equilibria of the system, so we can think of the bifurcation plot as showing how the attractor landscape changes across variation in k. We are looking for a correspondence between these equilibria and observed F0 extremum values in our pitch accent data. These equilibria, F^{equi}, are easily computed by solving, for each value of k, the roots of the polynomial on the right-hand side of the differential equation. As shown in Figure 11b, for the low values of k, below 1.19, there is a single equilibrium, which is an attractor at a high value of F. The color-matched (dark red) trajectories in Figure 11a rise to a low F0 maximum value relatively early. This is quite similar to the empirical H* trajectories presented in Section 3. As k increases in value, two more equilibria are born. This is sometimes called a *blue-sky bifurcation*, since two equilibria emerge out of the blue sky (without a triggering influence) but is more technically called an imperfect pitchfork bifurcation (Strogatz, 1994). The middle branch is a boundary, an unstable equilibrium, while the lowest branch is another stable equilibrium at a low value of F, as determined through a mathematical analysis (not included here; see Strogatz, 1994). Figure 11b also shows a horizontal black line at the fixed IC value = -.2 value. When $k <$ 1.19, the IC is below the only stable equilibrium, and the corresponding trajectories in Figure 11a simply rise towards that stable equilibrium, as there are no competing forces to influence those trajectories. For $1.19 < k < 2.54$, the IC is above the repellor boundary (the middle branch in Figure 11b) and accompanying trajectories in Figure 11a (light orange through yellow-green) are, therefore, repelled upwards. Early in that interval of $1.19 < k < 2.54$, the IC is relatively distant from the boundary (the middle branch in Figure 11b is below the IC), and therefore corresponding trajectories rise without sticking, whereas later in that same interval, the IC is quite close to the boundary, leading to sticking. When $k > 2.54$, the IC is below the boundary, if only slightly (the middle branch crosses the IC, trending upward), so trajectories are repelled downward.

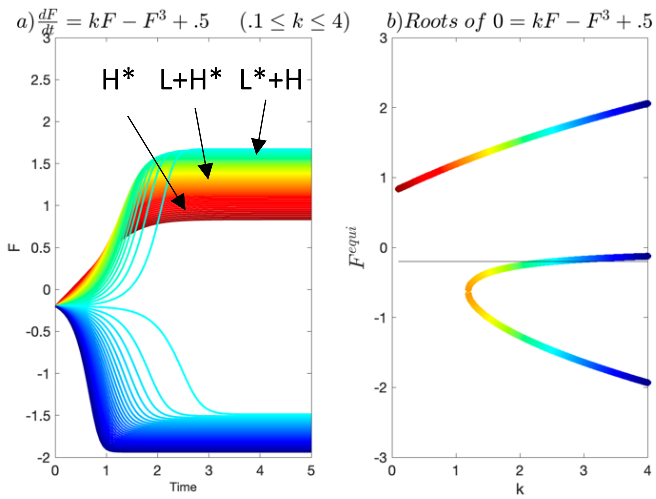

Figure 11. Model where IC is fixed at -.2, C is fixed at .5, while k varies. (a) Solutions (b) Bifurcation diagram, where F^{equi} are F values that make $\frac{dF}{dt} = 0$, so they are roots of the right-hand sides of the differential equation. Pitch accent labels are explained in the text.

We propose that different intervals along the range of k values generate F0 trajectories that correspond to different MAE pitch accent categories in the AM model, as follows. Typical F0 trajectories for the H* category correspond to the trajectories generated with k values in the first interval of $k < 1.19$, trajectories typical of L+H* correspond to k in the lower interval of the region $1.19 < k < 2.54$, while L*+H trajectories correspond to k in the higher intervals of the same k region, and L* trajectories correspond to the higher interval of $k > 2.54$. A truly remarkable property of this model is that the bifurcation, as an emergent property of the model, corresponds to what is a **categorical** distinction in AM between the monotonal (H*) and bitonal (L+H*, L*+H) pitch accents. Recall from the discussion in Section 3 (Table 3) that encoding this distinction in AM requires two constructs: contrastive level tones (H, L) and linearity (+). In the proposed dynamical model, the distinction between monotonal and bitonal rising is emergent, corresponding to two intervals in the range of k values: a lower interval where there is only a single attractor (stable equilibrium) at high positive values of F and a middle interval where there are two attractors, one positive and one negative. Establishing this correspondence between attractors in the dynamical model and the AM level tones (H, L), we see that the dynamical model generates one set of trajectories in a region where there is a sole attractor, H, and another set of trajectories in a region where there are two attractors, H and L. Said differently, just as the AM encoding of H* trajectories is monotonal and therefore absent the L feature, the corresponding behavior of the dynamical system exhibits only one branch, a high attractor. Moreover, like AM, the distinction marked by the bifurcation is categorical as well: for long stretches of k, small differences in a parameter lead to quantitative differences in the equilibria, whereas across the bifurcation value of k, small changes in k lead to qualitative differences in the equilibria.

This tight matching between AM and the proposed k-dynamic theory of pitch accents permits us to call the dynamical theory *phonological*. Like AM, the k-dynamic theory models the "discrete linguistics" of categorical contrast (Pierrehumbert and Pierrehumbert, 1990). In fact, there are two inter-related levels of discreteness in our dynamical model. The first, lower level is the discretization of the F-continuum by equilibria. The second, higher level, is the discretization of the k-continuum through bifurcation of equilibria. We consider this a phonological model because, although the model deals with F0 trajectories that vary continuously in F0 space and over time, the model generates equilibria as well as categorical changes in these equilibria. Goldsmith and Larson (1990), Goldsmith (1994), and Prince (1993) have already demonstrated how dynamical computation can derive discrete constructs, e.g., syllables, feet, and barriers to stress. Our proposal builds on this prior work in the attempt to bridge the discrete and continuous elements of intonation that characterize the traditional distinction between intonational phonology and its phonetic implementation. The IPB system in (1) computes F0 dynamics over a continuous time dimension, and in doing so goes beyond the work just cited, which model the dynamics of sonority and stress prominence over time that is already quantized in discrete units corresponding to segments (for syllable models) and syllables (for models of stress feet). As we have shown, discrete contrasts emerge from our model without prior quantization of the input, effectively capturing phonological contrasts between High and Low level tones and between monotonal and bitonal accents.

Looking beyond discreteness, the proposed dynamical system also effectively models within-category variation in F0 trajectories that is relegated to phonetic implementation in the AM model, though without recourse to a separate F0 implementation algorithm. For example, note how the H branch in Figure 11b rises, reflecting the scaling differences in the F0 maxima of rising accents noted in our empirical data (see Figures 2 and 5), a pattern that is also observed in other work on American English intonation (Arvaniti and Garding, 2007; Burdin et al, 2022). This variation in F0 peak scaling is what we describe earlier as the pattern of *rise-later-rise-higher,* which yields a kind of F0 prominence ranking among the rising accents: H* < L+H* < L*+H. Yet beyond our data, and the similar findings in Arvaniti and Garding (2007) and Burdin et al. (2022), this pattern linking F0 peaks, their temporal alignment, velocity and overall F0 span, is not discussed in prior work on MAE pitch accents, perhaps due to the more limited focus of most work on only one type of measurement, e.g., F0 range or F0 peak alignment. As such, we don't know the extent to which the pattern truly generalizes. It is therefore worth considering whether our dynamical model would still be appropriate in the case that no scaling differences are observed. We believe the answer is yes, since the earlier-later scale of latency can be decoupled from the higher-lower attractor scale, if the interval of k values that a language uses for linguistic functions is sufficiently narrow. Note especially in Figure 11b, where k values increase incrementally and uniformly, how in the cyan region, right before beginning of the L* interval, equal increments of increase in k leads to larger and larger delays in rise, with very small changes in the height of the F attractor. Therefore, a language can use this region to achieve large changes in latency (peak alignment), with little or no changes in the value of the attractor (the level tone).

4.2.3 The Fitzhugh-Nagumo dynamical system for MAE pitch accents

There is one behavior that the IPB system in (1) cannot accomplish, and that is generating a scooped trajectory that falls and then rises within the temporal interval of the pitch accent. We have observed this property in empirical F0 trajectories, especially those corresponding to the L*+H pitch accent, as shown in Figure 2 (top panel), and we have discussed this property in terms of velocity profiles shown in Figure 7. According to the IPB system in (1), an F0 trajectory will either rise or fall, but it is not possible for a trajectory to first fall and then rise. This is due to the repulsive threshold (middle branch in the bifurcation diagram in Figure 11b): Given a fixed IC, if k is below the critical point (k=1.19), F can only rise from the IC; for k above the critical point, when there are 2 stable equilibria, F will be repelled upwards if the IC is above the threshold and will be repelled downwards to the lower branch when the IC is below the threshold. This repulsive threshold dynamic is binary, generating rises or falls. Therefore, one more complexification of the dynamic needs to take place to account for L*+H scooping. It may seem that this is a small effect that can be ignored in a first account of pitch accent dynamics. However, we believe that an account of this fall-rise behavior is necessary for a more general dynamical theory of tonal behavior; after all, there are other languages that have a fall-rise as part of their tonal inventories, e.g., Mandarin Tone 3. We therefore develop a complexification of the basic IPB system that can account for scooping, but which is also motivated by building a general dynamical framework for tonal phenomena that goes beyond American English pitch accents.

There are two approaches that could be taken here, one nonautonomous and the other autonomous (Sorenson and Gafos, 2016). The nonautonomous approach would be to stipulate some extrinsic time-varying force that first pushes F towards a low equilibrium, then after a while that force changes polarity, pushing F upwards. This is a nonautonomous approach because the effect is artificially superimposed through this external time-varying force. The approach we take is autonomous as we agree with Sorenson and Gafos (2016) that explanations relying on system internal explanation are more powerful than ones that resort to extrinsic force sources. The latter lead to data-fitting rather than explanation, since such external forces can be multiplied, literally forcing F to yield whatever data is observed—the type of possibility that Pierrehumbert and Pierrehumbert (1990) warned against.

Phenomena in which a function first heads towards one value and then reverses direction to go towards another are well known in mathematical physics, biology, and neuroscience (Strogatz, 1994; Izhikevich, 2010). Indeed, in one of the first papers in the field now called Mathematical Neuroscience, "Mathematical Models of Threshold Phenomena in the Nerve Membrane", Fitzhugh (1955) presents an autonomous dynamical model that can handle phenomena like ours, where thresholds are far more complex than repellors. These models complexify the notion of *threshold* by adding a new level of system self-regulation. Besides the variable F, these accounts propose that there is another variable, which we will call I, that interacts dynamically with F in a simple way: I opposes the growth of F, but I is itself due to F.[11] Specifically, if I is positive, it inhibits F, i.e., it reduces the value of F. But if I is negative, it excites F, increasing it. What makes this account autonomous is that I is not an external force,

[11] In Section 5 we will offer one possible interpretation of what I is, but for now, it will simply be regarded as a necessary dynamical regulator of F.

but a variable that is itself regulated by F: if F is positive, I increases, and if F is negative, I decreases. The resulting system specifies two *coupled* differential equations for F and I, where the coupling specifies how each plays a role in regulating the other. Fitzhugh (1955) proposed a general type of system, now called the Fitzhugh-Nagumo[12] (F-N) system), in which F and I regulate each other as above. We have tailored their system to the demands of the MAE pitch accent system by making it non-oscillatory, while maintaining its insights about complex thresholds to produce this F-N System:

$$\frac{dF}{dt} = kF - F^3 - I + .5 \qquad \frac{dI}{dt} = .2F - .4I \tag{2}$$

4.2.4 An in-depth look at the F-N system

In this subsection we describe the meaning of the F-N system of differential equations, showing its solutions, and providing a graphical analysis of its implications. The goal here is to explain and illustrate, in some detail, how the system of differential equations in (2) generates the desired dynamical properties required to model our empirical data.

The differential equation for F is the same as for the IPB system in (1), except that another effect on F now comes from the -I term. This means that if I is positive, F is reduced, and if I is negative, F is boosted. The oppositional effect of I on F is, in turn, regulated by F through the positive .2F term in the differential equation for I. So, if F is positive, I increases, and if it's negative, I decreases. I also self-regulates by negative feedback due to the -.4I term. The specific values of coefficients in the second equation of (2) were chosen to make sure that the system does not oscillate, as required in most applications of the F-N system. The autonomy of this system is due to the interlocked nature of the two variables: Each of the two variables changes due to themselves and their relation with the other variable, creating a *self-organizing* system. The IC for I is .2, which is consistent with starting the system in an inhibited state, as we have previously assumed (recall that I is the inhibitor in this system, so a positive value promotes inhibition). The solutions of this system, and their velocity, are shown in Figure 12a,b. The trajectories generated by this system capture the distinctive shapes of the empirical trajectories shown in Figure 2 (top panel). Most importantly, some of the F trajectories (cyan) fall before rising, creating the characteristic scooped rise, as can be seen in how the corresponding velocity curves in Figure 12b start with negative velocity, then assume positive velocity (cyan). This panel also shows that for rises (trajectories with higher and higher k), peak velocities increase in value and occur with later delay (latency). Figure 12c also shows that as the value of k increases across the k-interval of the rising accents, the extreme value of F (the stable attractor) also increases, and then abruptly falls when k enters the k-interval for the falling accent. Also, peak velocity and time to peak velocity are highly consistent with the data.

[12] Nagumo, an electronics engineer, developed the same system at the same time to account for nonlinear electronic components.

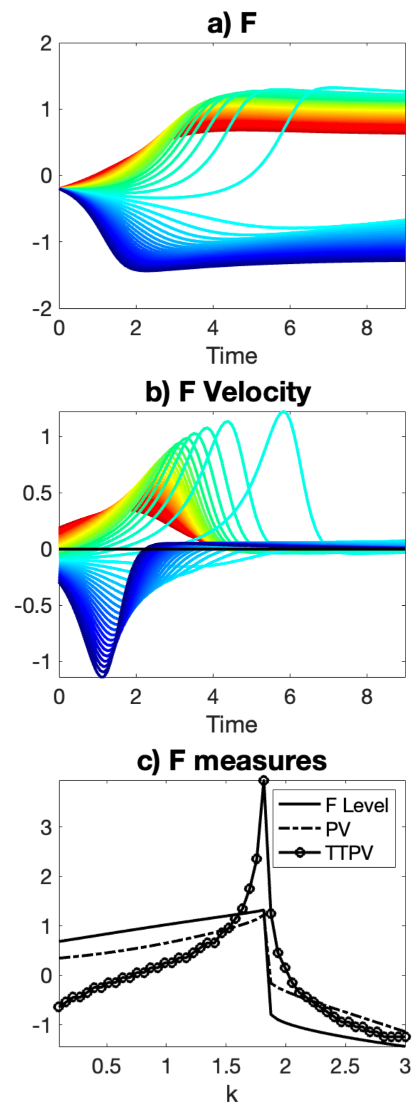

Figure 12. Final model, with 2 interacting variables: F and an inhibitor variable I. a) F Solutions; b) Velocities; c) Maxima and minima of F as k varies; Peak Velocity as a function of k; and Time to Peak Velocity (= latency) as a function of k.

Having observed that the F-N system exhibits behaviors needed to model the empirical behavior, we now turn to a dynamical explanation for these behaviors, and in particular, the potential for scooped trajectories. To do so, we need to find the equilibria for F and I, and for changes in these equilibria as k changes. This may seem to be more complex for 2 variables than 1, but the nullcline plot technique (Fitzhugh, 1955; Izhikevich, 2010) makes this task much easier. Figure 14 shows four such nullcline plots for 4 values of k. We will now gradually introduce the reader to the meaning of these plots, the curves on them, and how they explain the dynamical structure which we claim to underlie the MAE pitch accents. The axes represent specific values for F and I, and if these values are plugged into the right-hand side of (2) we get a value for the left-hand side $\left(\frac{dF}{dt}, \frac{dI}{dt}\right)$, which is drawn as an arrow at the (F,I) point. The arrow points to the next state of the dynamical system given the previous state. One can therefore start at any point (F,I) as the initial condition, and follow, graphically, the evolution of solutions (F(t),I(t)) by following from arrow to arrow from each time step to the next. In each panel of Figure 14, we show solutions emanating from the same IC at the red cross for four systems with different values of k and show the evolving solution by following the arrows. The direction of the arrow signifies the direction of change for F and I as follows:

↗	↘	↙	↖
$\frac{dF}{dt} > 0$ and $\frac{dI}{dt} > 0$	$\frac{dF}{dt} > 0$ and $\frac{dI}{dt} < 0$	$\frac{dF}{dt} < 0$ and $\frac{dI}{dt} < 0$	$\frac{dF}{dt} < 0$ and $\frac{dI}{dt} > 0$
F increases; I increases	F increases; I decreases	F decreases; I decreases	F decreases; I increases

Figure 13. Key to arrows on nullcline plots in Fig. 14, 18

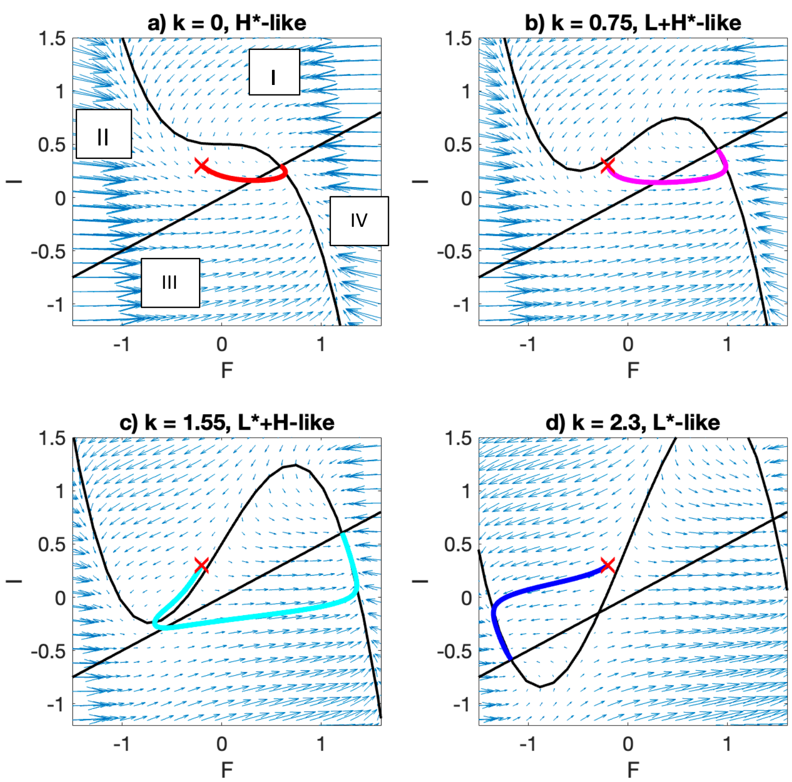

Figure 14. Nullcline plots for 4 k values corresponding to traditional AM labels: k = 0 is H-like, k = .75 is L+H*-like, k = 1.55 is L*+H-like, and k = 2.3 is L*-like. Each point in the (F,I) plane represents an IC. The arrow at each point expresses $\left(\frac{dF}{dt}, \frac{dI}{dt}\right)$, so for example, a rightward arrow expresses $\frac{dF}{dt} > 0$, and a downward arrow expresses $\frac{dI}{dt} < 0$. The black line expresses (F,I) that make $\frac{dI}{dt} = 0$, while the black cubic expresses $\frac{dF}{dt} = 0$. The red cross expresses the constant IC for all simulations for pitch accents, and the colored curves show (F(t),I(t)). Roman numerals in panel a label each discrete region discussed in the text.*

There are many possible (F,I), and accordingly, many arrows in the nullcline plots. A cursory look at these *vector fields* may lead one to believe that the dynamics for this system are incredibly complex due to the continuous variation in and large variety of vector magnitudes and directions. However, despite the enormous quantitative differences in arrows there is a small number of discrete behaviors in each panel. To see this qualitative simplicity, we also plot in each panel the black curve F^{equi} all of whose (F,I) points make $\frac{dF}{dt} = 0$. This curve is obtained by setting $\frac{dF}{dt}$ in Equation (2) to 0, then rearranging the top equation so that I is a cubic polynomial of F, hence the cubic shape of this curve. The straight line on the plots signifies I^{equi} points at which $\frac{dI}{dt} = 0$, and it's a line, since when $\frac{dI}{dt}$ is set to 0, the second equation of (2) can be rearranged so that I is a linear function of F. F^{equi} and I^{equi} are the *nullclines*, as the vectors on the nullcline have null magnitude in one direction. Of paramount importance to the explanations that follow is the fact that as nullclines approach each other, both F and I change less and less. Where they actually intersect are the combined (F,I) equilibria of the system. The F^{equi} curve and I^{equi} line intersect at one or more points depending on the value of k as can be seen in Figure 14. The F^{equi} curve and I^{equi} line divide the vector fields into a few uniformly behaving regions. This is most easily seen in the k = 0 plot (Figure 14a), where the and F^{equi} curve and I^{equi} line divide the (F,I) plane into 4 basic dynamical regions: I) above both curves, both F and I decrease; II) to the left of the cubic and above the line, F increases and I decreases; III) below the line and to the left of the cubic, both F and I increase; IV) below the line, and to the right of the cubic, F decreases and I increases. Within each region, there is quantitative variation, but all arrows point in one direction. Therefore, the infinite continuum of the (F,I) plane is divided into 4 types of *discrete* behaviors.

The increase in the linear term coefficient k in Equation (2) leads to a more and more pronounced positive-sloped linear portion of the cubic nullcline. As that positive-sloped portion gets longer and longer, the left half of F^{equi} goes lower and lower, coming increasingly close to I^{equi}, as can be seen if one scans from panel a to panel d of Figure 14. And as the shape of the cubic changes, there is a major qualitative shift. For k values below a certain bifurcation value (panels a-c), there is only one intersection of F^{equi} curve and I^{equi} line, and therefore one equilibrium, at a high value of F. For such values of k, wherever (F,I) starts, it will eventually get to that high equilibrium value. For k values above the critical point of k, there are 3 intersection points: an attractive high value, an attractive low, and a third intersection between them. We also note that the (F,I) plane is discretized into 6 qualitatively distinct regions when there are 3 equilibria. If the IC is to the right of the middle intersection, the high equilibrium will be achieved, and if it's to the left, then the low equilibrium will be achieved. Since we assume a fixed IC with negative F and positive I (marked by the red cross), then for k values below the critical value (i.e., the bifurcation), this model predicts rises such as for the pitch accents H* , L+H*, and L*+H, and a simple fall for k above the critical value.

Going from panel a to c in Figure 14, as the positive-slope valued portion of F^{equi} gets closer to I^{equi}, the magnitudes of the arrows get smaller, since greater proximity of F^{equi} and I^{equi} leads to reduction in the magnitudes of $\left(\frac{dF}{dt}, \frac{dI}{dt}\right)$. Therefore, as the value of k increases, but with only one high equilibrium (from panels a-c), the F trajectories get *slower* and *slower* in moving towards that high equilibrium, leading to greater and greater latency, which we associate with going from H* to L+H* to L*+H. Therefore, AM's notion of a linear sequence of an L before H in L+H* and L*+H, but not in H*, emerges from this model: when k is very small, for H*, F heads for the high equilibrium right away; as k gets larger, for L+H* and L*+H, the greater proximity of F^{equi} and I^{equi} leads to significant lingering in the low F region, before it escapes to the high equilibrium. The linear concatenation of L and H in the two bitonal accents L+H* and L*+H, we claim, emerges from this *k*-variation below the critical point.

What distinguishes panel a and b from panel c is that in the former, the IC is in region II (to the right of the cubic curve), and in this region F *monotonically* increases to the high equilibrium. In panel c, the IC is now in region I, where F decreases. And as I is positive, F decreases further, which further decreases I, until I is negative. And when that happens, I starts to boost F, therefore F moves to the high equilibrium. Therefore, the scooped shape of many rising trajectories in the L*+H category in our empirical data emerges from the dynamic at k values just below the critical point. Turning now to panel d, with an even higher value of k, the IC is again in region I, so F decreases, but there is now an attractor in the low F region, where F^{equi} intersects I^{equi}, and the F trajectory (F(t),I(t)) is attracted to this low equilibrium after the initial fall. This trajectory corresponds to L*. What this F-N model has added beyond the behavior of the IPB system in (1) is the possibility of an F trajectory that is briefly attracted to a lower F equilibrium before turning to the higher one—falling then rising. One of the predictions of this model is that the F0 minimum of L*+H is not as low as the minimum in L*.[13]

To summarize our observations from the nullcline plots, we find discreteness arising from two properties in this system: 1) the location of the IC with respect to equilibria; 2) the number of intersections. Together, these properties determine behaviors that correspond to qualitative, categorical distinctions among AM pitch accent categories, as well as finer-grained, quantitative distinctions among F0 trajectories as seen in Figure 12a.

Our observations from the nullcline plots in Figure 14 reflect only one set of initial values for F and I, so the question arises whether similar system behaviors would be observed under different initial conditions. Would we observe the same changes in the attractor landscape? We tested this and the findings show changing F-IC and I-IC results in system behaviors that are very similar to what is shown in Figure 14. The relatively stable behavior of the system reflects the fact that it is the presence of the IC within the 4 or 6 discrete regions, and not its precise placement in the region, that matters. That said, it is also possible that L*+H, for instance, could start with a lower F-IC than H*, as long as the IC's are in the same regions of the (F,I) plane we have assumed. Therefore, the theory we have proposed does not necessitate any particular or precise values for the F-IC and I-IC, nor does it require that all pitch accents start at the same value.

[13] This difference in F0 minima has not, to our knowledge, been explicitly addressed in prior work, and the comparison is difficult when taking coarticulatory effects from upcoming intonational features into account. For instance, our empirical data confirm a lower F0 minimum for L* compared to L*+H, but only in the L-L% edge tone context. We leave further testing of this prediction to future work.

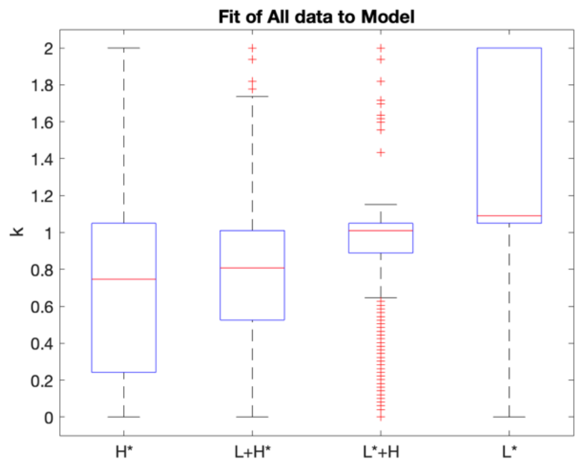

Figure 15. k's estimated from our empirical data, using the coupled system of differential equations for F and I.

4.3 Induction of k

So far, we have built a complex dynamical system molecule from small constituents, appropriately combined to generate the properties we discussed earlier. It is important to note that the F-N system of coupled differential equations in (2) has only 1 parameter, k, whose variation results in the *generation* of an entire scale of trajectories, some of which correspond to the F0 trajectories used for prominence signaling in MAE. An important question to ask is whether k can be induced from the data presented in Section 3. If so, this would lead to a very concise and inferable parameterization, with only one degree of freedom, of highly variable empirical F0 trajectories. To attempt this induction, we generated 1000 trajectories from this system of differential equations and for each, we calculated the velocity (the difference between two consecutive samples divided by *dt*). Then we computed velocity in the same way for each of the 12,743 of empirical trajectories. We then used the inner product to measure similarity of each empirical velocity trajectory to each of the 1000 theoretical velocity trajectories, and picked the k that leads to the highest inner product. Figure 15 shows the result of such induction, plotting the distribution of k values for each empirical F0 trajectory according to its pitch accent label (recall that these F0 trajectories are elicited in an intonation imitation paradigm; the labels identify the pitch accent of the model utterance presented on each trial). Elicited H* trajectories yield the lowest induced k values, though the distribution is notably large, indicating a lower internal consistency in the imitated production of H* compared to the other accent categories. (Recall from the discussion of Figure 2a that speakers appear to have neutralized the contrast between H* and L*+H in the edge tone context H-H%.) The k values of elicited L+H* trajectories overlap with those of H* but positioned higher in the range of the H* distribution. Elicited L*+H trajectories have a very narrow distribution with k values that go right up to the value that defines the IC boundary crossing. The distribution of induced k values for L*+H trajectories corresponds to the k values in the nullcline plots for L+H* and L* in Figure 14 (panels c, d).[14] Induced k values above the top of the L*+H range correspond only to elicited L*. Despite considerable variability, we see that the sole parameter of this model, k, can be extracted with good reliability from empirical F0 trajectories. Of course, our database was collected under a very specific experimental design eliciting imitated nuclear tunes, and so it remains to be seen in further work whether the induction algorithm we have outlined, or modifications thereof, is useful for automatic labeling of F0 trajectories elicited through other methods, or spontaneously produced. If so, this approach holds promise as it may eliminate the need for human labeling of F0 trajectories, which is a major bottleneck in the study of intonation.

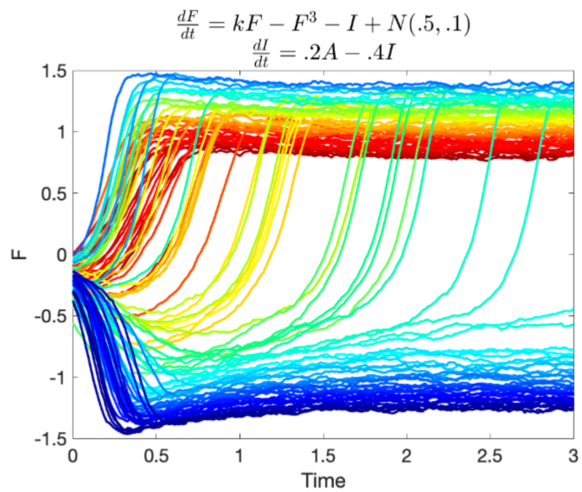

Figure 16. Solutions of the stochastic F, I system of differential equations, with noise in the initial conditions, as well as the F term. k-value is coded by color from lowest (red) to highest (blue).

4.4 A Stochastic Dynamical System

A body of experimental work investigating pitch accent production reports a striking degree of within-category variation in the F0 trajectories corresponding to the rising pitch accents (Chodroff & Cole, 2018, 2019; Im, Cole & Baumann, 2023; Ouyang & Kaiser, 2015; Ouyang, Spala, & Kaiser, 2017; Turnbull, 2017; see Cole, 2015 for general discussion)[15]. Our dynamical model addresses this, at least in part, by using a continuous variable k to parameterize

[14] The values of k selected to illustrate trajectories with different dynamics in Figure 14 do not exactly match the values of induced k values for our empirical F0 trajectories, though with additional iterations of the procedure we used to generate F trajectories from the F-N system, e.g., adjusting the constants in the model and the parsing of the tune into pitch accent and remainder, we could have generated k distributions that more precisely match the values selected in the illustrations of Figure 14. However, as it is not our primary goal to fit the model to this particular dataset with high precision, we chose not to optimize the algorithm for that purpose in this demonstration. We think it is even more remarkable that even without such data-fitting optimization, the F-N system in (2) generates F trajectories that capture the relationship between the pitch accents in our empirical data, both in terms of their *relative* k-distributions, and in terms of the empirical measures described in Section 3.

[15] We note here that there has been relatively less work examining F0 trajectories in production of the L* pitch accent. Also, related findings of phonetic variability in pitch accent production are reported for German and Italian (e.g., Baumann, Mertens, & Kalbertodt, 2019; Grice et al., 2017; Niebuhr et al., 2011; Röhr, Baumann, & Grice, 2022; see also Post, D'Imperio & Gussenhoven, 2007).

trajectories. This allows the theory to account for variability in F0 maxima and latency (i.e., peak height and alignment), but we do not claim that variation of these types cover all the types of variability observed. The deterministic dynamical systems we have introduced so far are idealizations of stochastic dynamical systems whose study has increased tremendously in the last several decades (Longtin, 2010), since most natural phenomena are highly variable. As an initial exploration we implemented a stochastic version of our model by adding gaussian noise to the right-hand side of each of the F differential equation, and setting the initial values of F and I (F-IC, I-IC) from gaussian distributions centered around -.2 for F and .3 for I. The results are shown in Figure 16. We now see some of the basic patterns we have seen before but with quite a bit of variation, which can be controlled by changing the standard deviation of the gaussian distributions. Observe that this model fairly closely replicates the dynamical properties of our non-stochastic implementation, in particular, the relationship between the F0 latency and maxima of the rising trajectories. We note that in very few instances, when a particular pitch accent is expected, a trajectory can be generated that is quite different from those expected for that accent, as can be seen in an L* tone being realized as a highly unexpected rising accent (the few rising blue curves in Figure 16). Current understanding of the extent and nature of stochastic variability does not allow us to insightfully determine the noise probability distribution and width, but this model is advanced to show that this framework is highly compatible with stochastic phonological models (Gafos and Benus, 2006).

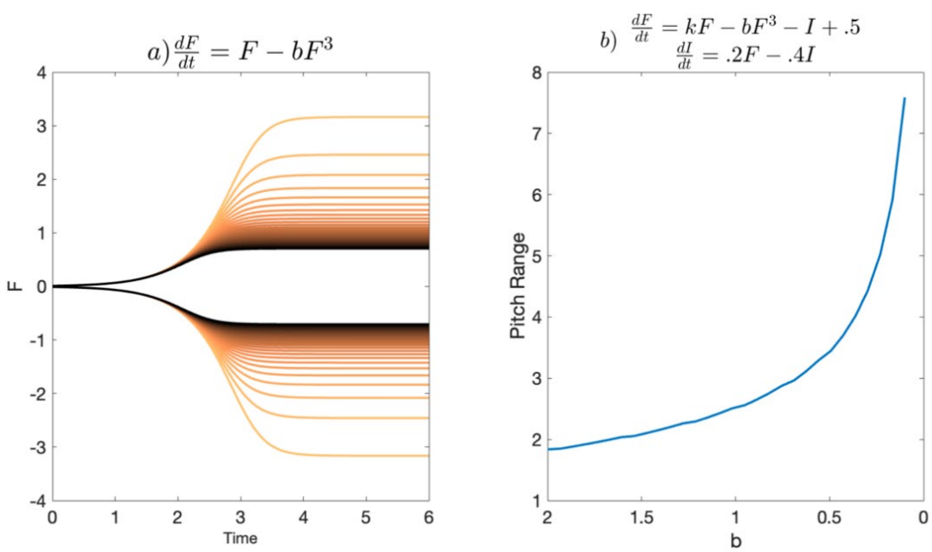

Figure 17. Pitch range variation introduced by the coefficient b of the cubic term. a) Simplest system with only linear positive feedback and negative cubic feedback—the narrowest F0 range is for the largest cubic b; b) Pitch range = highest equilibrium – lowest equilibrium, obtained across the k-range, as a function of b for the full coupled (F,I) differential equations.

4.5 Scale

Pitch range has often been observed to vary according to several factors, and the model we have presented is indeed scalable. The cubic negative feedback loop is responsible for stabilizing rising and falling trajectories at the high and low attractors, corresponding to High and Low tone targets, so it should not be surprising that this term can be modified to include scale. Adding a positive parameter b on the cubic to obtain the differential equation in (3) adds scale.

$$\frac{dF}{dt} = F - bF^3 \tag{3}$$

Figure 17a shows the F trajectories generated by (3) with ICs of .01 (for trajectories above 0) and -.01 (for trajectories below 0), but where b varies from .1 (light copper) to 2 (dark copper). Observe that as the cubic parameter changes, the scale changes. This is also true for the more complex model F-N model proposed here for the MAE pitch accent system. Figure 17b plots the F range (the distance between the top and bottom range, as shown in both branches of Figure 17a) for increasing values of b. Here it is important to distinguish between two concepts of scale. Earlier, in relation to the IPB system in (1) we discussed variation in the scaling of the high F equilibrium resulting from variation in k (see Figure 11). We think of this as a pitch accent-intrinsic scale, since the F0 maximum, or accentual peak, appears to increase in scale across different rising pitch accents, from H* to L*+H. The scale concept we now discuss is an extrinsic scale that modifies the whole pitch range. It remains to future research to examine how this extrinsic scale is dynamically modified in phenomena such as downstep.

5. Discussion

5.1 Emergent dynamical properties of the proposed model

The most specific goal for this work has been to provide a new, unified model of the phonetics and phonology of MAE pitch accents. Beyond that goal, our work also has theoretical import. We have introduced a dynamical theory of intonation from which the fundamental constructs of AM emerge, rather than being pre-specified, thereby addressing the critique offered by Pierrehumbert and Pierrehumbert (1990) for previously proposed dynamical models of speech production. The full model, accounting for all the measured empirical properties in our dataset, is the stochastic version of our model (2), repeated here as (4):

$$\begin{aligned}\frac{dF}{dt} &= kF - F^3 - I + N(.5, .1)\\ \frac{dI}{dt} &= .2F - .4I + N(0, .1)\end{aligned} \tag{4}$$

In developing this model, we adopted a combinatorial/hierarchical approach that allows us to claim that this model is *minimal*, as we started with the simplest differential equations, and built up, term by term to produce a minimally complex system that accounts for the generalizations from our data and other published data. The smallest nucleus of this system, $\frac{dF}{dt} = kF$, a positive feedback system, already accounts for the distinction between rising and falling F0 configurations. The addition of the negative feedback cubic term stabilizes these trajectories to level out at attractor states, the F equilibria, that correspond to the level tones, Low and High. The addition of a constant, the mean of the distribution N(.5,.1), has two related consequences: 1) it captures variation in latency, differentiating rises with early vs. late onsets, thereby eliminating the need for the distinct representational devices used in AM to encode linear sequences of tones (+) and the anchoring of one tone in a bitonal accent to a phonological landmark (*); 2) it captures the *rise-later-rise-higher* generalization seen in our data and observed elsewhere in the literature. The addition of the inhibitor, I, with its excitation-inhibition relation to F, allows for the possibility of trajectories that fall before rising, important for capturing the scooped shape of the late-aligned accent L*+H in the MAE system. Introduction of the dynamical noise sources into the dynamics and the initial conditions allows for variation in F0 trajectories as discussed in Cole (2015) and observed in many (if not all) prior studies.

We have shown, through the proposed model, that dynamical computation can generate discrete qualitative behaviors attributed to the MAE pitch accent system in AM analyses. We take this to constitute an upgrade in Task Dynamics, because our equations (unlike those of AP) do *not* have specific constants that specify that a Low or High value is to be achieved, that there should be a specific amount of delay, or that a particular high value of F0 should be preceded or not preceded by a low value. All these aspects are predicted by the interaction between terms in the dynamical equations parameterized by one interaction or *kinetic* parameter k that weighs the linear term. Pierrehumbert and Pierrehumbert (1990) conclude their article with "In 'discrete linguistics' we skip the dynamical middleman and go directly to trying to understand the behavior in terms of discrete representations obtained from observing the phenomenon itself (language produced by speakers). If any of these phenomena are ever to be accounted for on another level by a continuous dynamic representation, knowledge of the discrete dynamics will no doubt prove an essential clue to the reconstruction" (p. 476). AM is a theory of the "discrete linguistics" of pitch accent, and work in the AM framework constitutes a very important and useful step in the exploration of intonation in speech production, but we argue that skipping the dynamical middleman is no longer necessary. Because, as we hope to have shown, *basic* dynamical systems analysis is actually sufficient to allow discrete constructs to emerge, rather than be stipulated.

Yet some may wonder "why bother"? If discrete linguistics is a sufficient language for describing linguistic entities, perhaps knowing the dynamical origins is a nice *implementational* addition, but not *necessary* for describing language. We disagree. Knowing the dynamical theory from which the constructs emerge is essential for understanding the relation between pitch accents, and possibly other tonal constructs, of the languages of the world, and for describing the many possible bases of variability within and between speakers, dialects, and languages. The dynamical system, as a model of state change over time, is what takes the study of intonation from a set of observations into a predictive theory that describes a complex set of observations with as few dynamical atoms and parameters as possible. Furthermore, we may gain new insights into intonational systems by considering how dynamic properties vary in relation to linguistic function. For instance, our dynamical model of MAE pitch accents captures variation among the rising accents (H*, L+H*, L*+H) through variation in the free parameter k. There is a systematic ordering of these accents in terms of their empirical measures of F0—specifically, their velocity, latency, and span—which is mirrored by the same ordering of the k values associated with each accent category: H* < L+H* < L*+H, as shown in Figure 6.

The relationship between variation in F0 dynamics and variation in k values across the rising accents suggests that k may function to encode prosodic prominence. This interpretation is further strengthened by research on prominence perception showing that, when listeners rate words for perceived prominence, the likelihood of a prominence rating varies according to the pitch accent status of the word, with the same ranking among the rising accents (Cole, et al., 2019; Im, Baumann, & Cole, 2023). Yet, observing this relationship between prominence and k for the rising accents leaves us to wonder about the status of L*, which in our model is associated with the highest k values. On one hand, work on intonational meaning (e.g., Büring, 2016; Hirschberg, 2006; Hobbs 1990; Pierrehumbert & Hirschberg, 1990) describes L* as having low informational value: It is variously described as encoding "givenness", or the absence of predication or assertion, e.g., for words that are referentially or lexically salient from prior context. This suggests that L* marks words with lower informational prominence vis-à-vis the rising accents. On the other hand, the same prominence rating studies of MAE, cited above, that show a gradient likelihood of prominence rating across the rising accents from H* to L+H*, also show substantial variation in prominence rating for L* words, which are similar to H* in terms of their likelihood to be perceived as prominent. Thus, while L* is not at the top of the rating scale of perceived prominence, it is also not at the bottom (Cole et al., 2019; Im et al., 2023).[16] It seems, then, that if even if L* is not informationally prominent, it nonetheless registers as prominent for MAE listeners. Putting these observations together, it appears that while k relates to differences in information status and perceived prominence for the rising accents, it does not track the status of L* on either of those scales. An alternative may be to think about k in relation to markedness, in which case implementing an L* accent on a prominent word with a drop in pitch may be considered marked, in that it may be perceptually salient as distinct from the alternative of leaving the word unaccented. We leave this and other challenging questions about the linguistic functions of k for future research.

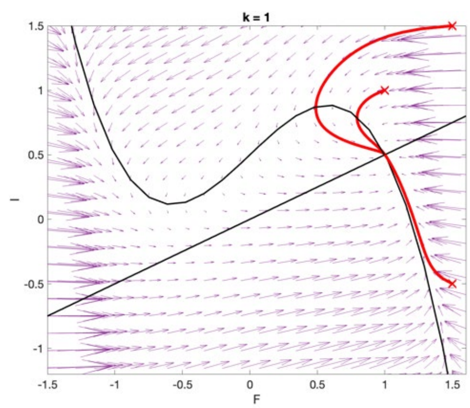

Figure 18 Possible falls from different positive F's.

[16] The perceptual prominence rating reported for L* in these studies of MAE differs from findings reported for German L* words, from studies using the same or similar prominence rating task (Baumann & Winter, 2018; Baumann & Röhr, 2015). German listeners preferentially perceive L* words as not prominent. To understand whether and how this difference in perceived prominence between MAE and German relates to empirical measures of the pitch accents (as in Table 2) in both languages would require a dynamical model of German pitch accents along the lines presented here for MAE.

5.2 Limitations of the proposed model and future extensions

To extend the proposed dynamic theory beyond the current focus on the analysis of MAE pitch accents to examine pitch accent systems of other dialects and languages, further research is required to examine the values of k and the scaling variable b (introduced for the cubic term in equation (3) above) that generate the patterns of F0 trajectories used in those languages. Taking a dynamic theory approach allows the possibility, for instance, of comparing English and Spanish L*+H with each other, without recourse to a representational level between discrete phonology and continuous phonetics (c.f., Hualde and Prieto, 2016; Ladd, 2008b) by hypothesizing that different languages choose different regions of k-space. Our focus in the present work is admittedly narrow. First, looking only at MAE, we have restricted our analysis to the three rising pitch accents of the AM model, augmented with the sole low-falling accent, which leaves out the high-falling accent H+!H* and the scaling-related phenomenon of downstep. The F-N model developed here is in fact capable of generating falling F0 trajectories, which is illustrated in the nullcline plot in Figure 18, where leftward F trajectories occur at different positive values of F-IC, with k=1. For all three of the trajectories shown in red, IC-F is higher than F^{equi}, the high attractor, so F falls (a leftward shift on the x-axis). Note that the H+!H* tone is considered to start by first raising the baseline IC, before the High equilibrium is achieved. These examples also show different types of system behavior with trajectories that only fall and those that fall and then rise. Research is necessary to classify falls in MAE and other languages in terms of where they fall from and how they fall, but the possibility of such F0 trajectories is already present in the proposed F-N system.

Second, we have not presented a dynamical model of the F0 trajectories arising from edge tone context (phrase accents and boundary tones in the AM model: H-H%, L-H%, H-L%, L-L%). We believe that the well-studied shapes of these combinations can be generated using the framework we have proposed—e.g., the F0 trajectory for L-H% is similar to that of L*+H, a rise that starts with a brief fall, and this is one reason we have expended substantial effort to show how this kind of trajectory arises in our dynamical system. Our ongoing work aims to extend the dynamical approach to model these phenomena, including edge tones and their coarticulatory interactions with pitch accents. We believe that the dynamical building blocks provided here have potential for developing a more general cross-linguistic theory of intonation, and that a similar approach may serve to develop a dynamical theory of lexical tone.

A further goal of our ongoing work concerns tonal systems in general. Some tonal systems have both High and Low tones, and we believe that k and b variation in our final system could provide a foundation for these systems as they do for the MAE pitch accent system. However, some tonal systems, such as many Bantu languages (Hyman, 2017; Goldsmith, 1976; Odden and Bickmore, 2014), revolve around H only, which from a dynamical systems perspective would require only a single equilibrium. We aim to show in future work that a system based on a quadratic, instead of a cubic, a refinement of the very first differential equation: $\frac{dF}{dt} = F - F^2$, will show only one H equilibrium. These are of course promissory notes at this stage, and it remains for further work by us, and we hope others, to develop the dynamical theory approach to tonal behavior.

5.3 Looking ahead: Unifying the phonology and physiology of intonation

In the article "On Distinctive Features and their Articulatory Implementation", Halle (1983) presents a theory of how agonist-antagonist pairs of muscle groups of the tongue, interconnected via excitation and inhibition, can implement the concepts of feature theory for vowels. He also argues that this excitation-inhibition "circuitry" can explain measurements on muscular activation during vowel production (Alfonso et al., 1982). This work is extremely interesting as it shows the abstract similarity of a cognitive theory of phonological opposition and a motor theory of interactive muscular opposition. Both systems seem to follow the same *oppositional logic.* For Halle (1983), we believe, that similarity of phonological and motor systems is taken to be just a surface similarity, as the design principles of language such as feature organization, and laws governing speech behavior such as motor circuits, are distinct. This can be seen in the very title, where the circuits are *implementational*. This view carries over in much of phonology, including intonational phonology, where phonetics provides bodily implementation of phonological patterns of the mind.

The similarity pointed out by Halle (1983) between phonological and physiological systems, namely, that operations in both domains are governed by relationships of opposition (distinctive features in phonology; excitatory-inhibitory forces in motor control), can be viewed simply as a coincidental surface similarity between the systems. However, we believe it is possible that the similarity is a signature of the computational architecture of linguistic action. Indeed, the F-N system presented in this paper has a straightforward interpretation as the law of the interaction of two sets of muscles. One set of muscles are the agonists F that set F0 in upward motion, via tensing of the vocal folds, and the other is a set of agonists I that oppose vocal fold tensing. Note that the F-N equations emerged in mathematical neuroscience, and have been used to describe motor systems, and therefore it is conceivable that the motor systems of the larynx that govern vocal fold tension could work in a way consistent with the F-N system in (2). We suggest this only as a conjecture at this stage. Taking this idea a step further, a phonological theory of vowel systems that is defined using the F-N architecture in (2) would provide a unified architecture for studying vocalic *and* intonational aspects of speech behavior.

One argument against such a unified theory of phonological and physiological systems, rests on the observation that linguistic systems differ across communities of practice, e.g., in the vocalic or prosodic realms, even though all humans have roughly the same physiology -- vocal tracts and larynges. Does not a unified theory of physiological-phonological structure predict one language corresponding to the one muscular architecture shared by all humans? The answer is no. What a system like the F-N system provides is a basic dynamical architecture, or set of principles (Goldsmith, 1994), that may be *parameterized* differently in different languages through choice of different interactional parameters like k and b. The regions of these parameters (or their distributions within set regions) that are available for association with distinctions in (pragmatic or lexical) meaning would constitute a phonological choice of a system for a linguistic community of practice. Here we must acknowledge that a great deal of further work on the tone and intonation systems of other languages may call for further complexification of the F-N system. For instance, a tone or intonation system with no phonological opposition between high and low tones, as has been proposed for some Bantu languages (Hyman, 2017; Goldsmith, 1976; Odden and Bickmore, 2014), can be represented by exclusion of the cubic term by setting b to 0, and replacing it by a quadratic term. The dynamical architecture required for such a system could be defined as in (5), replacing the model in (2) proposed here for MAE:

$$\begin{aligned} \frac{dF}{dt} &= kF - bF^3 - qF^2 + .5 \\ \frac{dI}{dt} &= .2F - .4I \end{aligned} \tag{5}$$

Further examination of systems that are not organized around pitch accents (e.g., Korean, Hindi, French; see Jun, 2005, 2014) could reveal an even more intricate set of parametrized principles of linguistic action.

5.4 Correspondence with dynamical model of intonation planning

A surprising aspect of the work we have presented is that there is a very precise mathematical relation between our penultimate model, given in (1), and the model of intonation planning proposed by Roessig, Mücke, and Grice (2019) and Roessig (2021). If the right-hand side of the differential equation (1) is integrated, and multiplied by -1, we would obtain a double potential, where k variation changes the depth of the potentials, a model that is equivalent to that in Roessig et al. (2019). The approach taken in that work is, of course, based on a sequence of models starting with Haken et al. (1985) showing how producing one behavior vs. another, perceiving one percept vs. another, or intending one thing vs. another can be modeled by varying the depth of a double-well potential to make one behavior more stable and therefore more likely by lowering its potential. These theories of *intentional dynamics* (or, planning theories) are quite deep and have influenced many other theoretical works in speech production (Tilsen, 2019), but they are fundamentally different from the dynamical approach taken here. Their focus is the planning and selection of behaviors or intention, not about the detailed temporal unfolding of produced behaviors, aspects of what is usually termed *execution*, as we have shown in this paper. A long line of anti-dualist research in cognition (e.g., Grossberg, 1973; Fowler, 1985) has suggested that planning and execution are made of the same cloth. What our work points to is that this cloth is encoded in the dynamical framework we have provided. And while planning and execution are usually regarded as different behaviors, they may involve similar computational organizing structures (Iskarous and Pouplier, 2022). This is a compelling perspective that we leave for future research.

Acknowledgments

This work was funded by NSF grants PAC-1246750 to the first author and BCS-1944773 to the second author. We would like to thank Dani Byrd, John Goldsmith, Matt Goldrick, Louis Goldstein, Jonathan Harrington, Keith Johnson, Mark Liberman, Marianne Pouplier, and Alessandro Vietti for valuable discussion. All errors are of course our own.

References

Abraham, R. and Shaw, D.C. (1992). Dynamics: The Geometry of Behavior. Aerial Press.

Alfonso, P. J., K. Honda, T. Baer, and K. S. Harris. (1982). 'Multi-Channel Study of Tongue EMG during Vowel Production,' *Paper presented at the 103rd Meeting of the Acoustical Society of America,* April 25-30, 1982.

Andronov, A. A., Vitt, A.A., and Khaikn, S.E. (1966). *Theory of Oscillators.* Dover Books.

Arvaniti, A. (2011). The representation of intonation. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *Blackwell Companion to Phonology* (pp. 757–780). Oxford, UK: Oxford University Press.

Arvaniti, A. 2016. Analytical decisions in intonation research and the role of representations: Lessons from Romani. *Laboratory Phonology*, 7(2), 6. DOI:http://dx.doi.org/10.5334/labphon.14

Arvaniti, A. 2019. Crosslinguistic variation, phonetic variability, and the formation of categories in intonation. *Proc. International Congress of Phonetic Sciences 2019*. https://assta.org/proceedings/ICPhS2019

Arvaniti, A. and Garding, G. (2007). Dialectal variation in the rising accents of American English. *Laboratory Phonology*, 9, 547-546.

Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, *3*(2012), 1–49. https://doi.org/10.1515/lp-2012-0017

Baumann, S., Mertens, J., & Kalbertodt, J. (2019). Informativeness and speaking style affect the realization of nuclear and prenuclear accents in German. In *Proc. of the International Congress of Phonetic Sciences*, Melbourne, Australia, pp. 1580-1584.

Baumann, S., & Röhr, C. (2015). The perceptual prominence of pitch accent types in German. *Proc. of the International Congress of Phonetic Sciences.* Glasgow, UK.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, *70*, 20–38. https://doi.org/10.1016/j.wocn.2018.05.004

Beckman, M. E., & Ayers, G. E. (1997). Guidelines for ToBI labelling, version 3.0. *The Ohio State University Research Foundation,* 1–43.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Phonology above the foot Intonational structure in Japanese and English. *Phonology Yearbook*, *3,* 255–309.

Bolinger, D.L. (1951) Intonation: Levels Versus Configurations, *Word*, 7:3, 199-210, DOI: 10.1080/00437956.1951.11659405

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201-251.

Bruce, Gösta. 1977. Accentuation and timing in Swedish. *Folia Linguistica, 17*(1-4), 1983, pp. 221-238. https://doi.org/10.1515/flin.1983.17.1-4.221

Burdin, R. S., Holliday, N. R., & Reed, P. E. (2022). American English pitch accents in variation: Pushing the boundaries of mainstream American English-ToBI conventions. *Journal of Phonetics*, *94*, 101163.

Büring, D. (2016). *Intonation and Meaning*. Oxford University Press.

Byrd, D., & Krivokapić, J. (2021). Cracking Prosody in Articulatory Phonology. *Annual Review of Linguistics, 7*, 31-53.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*, 149-180.

Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, *40*(2), 329–349. https://doi.org/10.1016/j.wocn.2011.12.001

Cangemi, F and Grice, M. 2016. The Importance of a Distributional Approach to Categoriality in Autosegmental-Metrical Accounts of Intonation. *Laboratory Phonology, 7*(1): 9, pp. 1–20, DOI: http://dx.doi.org/10.5334/labphon.28
Chodroff, E. R., & Cole, J. (2018). Information structure, affect, and prenuclear prominence in American English. In *Proceedings of INTERSPEECH 2018* (pp. 1848-1852). International Speech Communication Association.
Chodroff, E. R., & Cole, J. (2019). The phonological and phonetic encoding of information status in American English nuclear accents. In *Proceedings of the 19th International Congress of Phonetic Sciences*.
Cohen, J. (1988). *Statistical Power analysis for the Behavioral Sciences.* Routlage.
Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, *30*(1-2), 1-31.
Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, *75*, 113-147.
Cole, J., Steffman, J., & Tilsen, S. (2022). Shape matters: Machine classification and listeners' perceptual discrimination of American English intonational tunes. *Proceedings of Speech Prosody 2022*, 23-26.
Cole, J., Steffman, J., Shattuck-Hufnagel, S. & Tilsen, S., (2023) "Hierarchical distinctions in the production and perception of nuclear tunes in American English", *Laboratory Phonology* 14(1). doi: https://doi.org/10.16995/labphon.9437
Dilley LC, Ladd DR, Schepman A. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. Journal of Phonetics. 33, 115-9.
D'Imperio, M. 2000. The role of perception in defining tonal targets and their alignment. PhD thesis, Ohio State University.
Fitzhugh, R. (1955). Mathematical Models of Threshold Phenomena in the Nerve Membrane. *Bulletin of Mathematical Biophysics*, 17, 257-278.
Fowler, C. (1985). Current perspectives on language and speech production: a critical overview. In R. Daniloff (Ed.), *Speech science: recent advances* (pp. 193-278). San Diego, CA: College Hill Press.
Fowler, C., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language Production. Volume1: Speech and Talk* (pp. 373-420). London: Academic Press.
Fujisaki, H. (1997). Prosody, Models, and Spontaneous Speech. In: Sagisaka, Y., Campbell, N., Higuchi, N. (eds) *Computing Prosody*. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-2258-3_3
Gafos, A. I., and Benus, S. (2006). Dynamics of phonological cognition. *Cogn. Sci.* 30, 905–943. doi: 10.1207/s15516709cog0000_80
Gao, M. (2009). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Chinese Journal of Phonetics, 2*, 43-50.
Goldsmith, J. A. (1976). *Autosegmental phonology.* Doctoral dissertation, Massachusetts Institute of Technology.
Goldsmith, J. (1974). *English as a tone language*. Distributed in 1981 in D. Goyvaerts (ed). *Phonology in the 1980's. Story-Scientia.*

Goldsmith, J. (1994). A dynamic computational theory of accent systems. In *Perspectives in Phonology*, ed. Jennifer Cole and Charles Kisseberth, 1‑28. Stanford, CA: Center for the Study of Language and Information.
Goldsmith, J., and Larson, G. (1990). Local modeling and syllabification. In *Papers from the 26th Annual Regional Meeting of the Chicago Linguistic Society*, part 2, ed. K. M. Ziolkowski, K. Deaton, and M. Noske, 129‑142. Chicago: Chicago Linguistic Society.
Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107. https://doi.org/10.1016/j.wocn.2017.03.003
Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213–257.
Haken H., Kelso J., & Bunz H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51:347‑356. [PubMed: 3978150].
Halle, M. (1983). On Distinctive Features and their Articulatory Implementation. *Natural Language and Linguistic Theory*, 1, 91-105.
Haugen, Einar. 1949. Phoneme or prosodeme*? Language, 25*: 278-282.
Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word initial clusters - the case of 'impure s'. *Phonology, 30*, 1-25.
Hirschberg, J. (2006). Pragmatics and intonation. *The handbook of pragmatics*, 515-537.
Hirst, D., Di Cristo, A., & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In *Prosody: Theory and experiment: Studies presented to Gösta Bruce* (pp. 51-87). Springer Netherlands.
Hobbs, J. (1990). The Pierrehumbert-Hirschberg theory of intonational meaning made simple: Comments on Pierrehumbert and Hirschberg. *Intentions in communication*, 313-323.
Hualde, J. I. and Prieto, P., (2016) “Towards an International Prosodic Alphabet (IPrA)”, *Laboratory Phonology*, 7, 1-25.
Hyman, L. M. (2017). Bantu Tone Overview. *UC Berkeley PhonLab Annual Report*, 13. http://dx.doi.org/10.5070/P7131040751
Im, S., Cole, J., & Baumann, S. (2023). Standing out in context: Prominence in the production and perception of public speech. *Laboratory Phonology 14*(1), 1-62. https://doi.org/10.16995/labphon.6417
Iskarous, K. (2017). The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics,* 64, 8-20.
Iskarous, K. (2019). The morphogenesis of speech gestures: From local computations to global patterns. Front Psychol. 2019;10:2395. PubMed Central PMCID: PMC6861444
Iskarous, K. and Goldstein, L. (2018). The dynamics of prominence profiles: from local computation to global patterns. *Shaping Phonology* (Editors: Diane Brentari and Jackson Lee). University of Chicago Press.
Iskarous, K., Pouplier, M. (2022). Advancements of phonetics in the 21st century: a critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics,* 2, 95, 101195.
Izhikevich, E. (2010). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting.* MIT.

Jun, S. A. (Ed.). (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, S. A. (Ed.). (2014). *Prosodic typology II: the phonology of intonation and phrasing*. Oxford: Oxford University Press.

Karlin, R. and Tilsen, S. (2014). The articulatory tone-bearing unit: Gestural coordination of lexical tone in Thai, *Proceedings of Meetings on Acoustics*, 22, 1.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics, 55*, 149-181.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics, 44*, 62-82.

Kawahara, H., Cheveigné, A. D., Banno, H., Takahashi, T. & Irino, T. (2005). Nearly defect-free f0 trajectory extraction for expressive speech modifications based on STRAIGHT. In *Ninth European Conference on Speech Communication and Technology.*

Krivokapić, J. (2020). Prosody in Articulatory Phonology. In S. Shattuck-Hufnagel & J. Barnes (Eds.), *Prosodic Theory and Practice*. Cambridge, MA: MIT Press.

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Phil. Trans. R. Soc. B*.

Krivokapić, J., Styler, W., & Parrell, B. (2020). Pause postures: The relationship between articulation and cognitive processes during pauses. *Journal of Phonetics, 79*, 100953.

Ladd, D.R. 1983. Phonological features of intonational peaks. *Language, 59*: 517-39.

Ladd, D. R. (2008a). *Intonational phonology,* 2nd edition. Cambridge, UK: Cambridge University Press. 1st edition published 1996.

Ladd, D. R. (2008b). Review of Sun-Ah Jun (ed.)(2005). *Prosodic typology: the phonology of intonation and phrasing*, 372-376.

Ladd, D. R. (2022). The trouble with ToBI. In Barnes, J. & Shattuck-Hufnagel, S. (Eds.). *Prosodic theory and practice*, 247-258.

Ladd, D. R., & Schepman, A. (2003). ‘‘Sagging transitions’’ between high pitch accents in English: Experimental evidence. *Journal of Phonetics, 31*(1), 81–112.

Longtin, A. (2010). Stochastic dynamical systems. *Scholarpedia*, 5(4): 1619.

Maeda, S. (1976). *A Characterization of American English Intonation*. Unpublished Ph.d. Dissertation, MIT.

Niebuhr, O. 2007. The signalling of German rising-falling intonation categories - The interplay of synchronization, shape, and height. *Phonetica, 64,* 174-193.

Niebuhr, O., D'Imperio, M., Fivela, B. G., & Cangemi, F. (2011). Are There" Shapers" and" Aligners"? Individual Differences in Signalling Pitch Accent Category. *Proc. International Congress of Phonetic Sciences,* pp. 120-123.

O'Connor, J. D. & G. F. Arnold. 1973. *Intonation of colloquial English.* London: Longman.

Odden D.A., Bickmore, L. Melodic tone in Bantu: overview. *Africana Linguistica 20*, pp. 3-13. doi: https://doi.org/10.3406/aflin.2014.1021.

O’Shaughnessy, D. (1976). Mode*lling Fundamental Frequency, and its Relationship to Syntax, Semantics, and Phonetics*. Unpublished Ph.d. Dissertation, MIT.

Ouyang, I, and Kaiser, E. (2015). Prosody and information in a tone language: an investigation of Mandarin Chinese. *Language, Cognition, and Neuroscience,* 30, 57-72.

Ouyang, I. C., Spala, S., & Kaiser, E. (2017). Speakers' rapidly-updated expectations influence prosodic realization of information structure *Proceedings of the Linguistic Society of America*, *2*(51), 1–10.

Perrier, P., Abry, C., & Keller, E. (1988). Vers une modelisation de la langue. *Bulletin de la Communication,* 2, 181, 45-63.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* [PhD thesis]. Massachusetts Institute of Technology.

Pierrehumbert, J. B. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America*, *70*(4), 985–995. https://doi.org/10.1121/1.387033

Pierrehumbert, J.B., & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse.

Pierrehumbert, J.B. & Pierrehumbert, R.T. (1990). On attributing grammars to dynamical systems. *Journal of Phonetics*, *18*, 465-477.

Pierrehumbert, J. B. & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, *46*(4), 181-196.

Pike, K. L. (1945). *The intonation of American English.* Ann Arbor: University of Michigan Press.

Post, B., D'Imperio, M. & Gussenhoven, C. (2007). Fine phonetic detail and intonational meaning. *Proc. of the International Congress of Phonetic Sciences*, Saarbruecken, Germany, 191-196.

Prince, Alan. 1993. *In Defense of the Number i: Anatomy of a Linear Dynamical Mode of Linguistic Generalizations.* Technical Report 1. Piscataway, NJ: Rutgers University Center for Cognitive Science.

Prince, Alan (2007). The Pursuit of Theory. In Paul de Lacy, ed., *Cambridge Handbook of Phonology*. Cambridge University Press.

Prom-on, S., Xu, Y., Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *JASA,* 125, 405-424.

Roessig, S. (2021). *Categoriality and continuity in prosodic prominence*. Language Science Press.

Roessig, S., Mücke, D., Grice. M. (2019). The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLOS ONE,* 15, e0231221.

Röhr, Christine T., Stefan Baumann & Martine Grice (2022). The influence of expectations on tonal cues to prominence. *Journal of Phonetics, 94,* 101174. https://doi.org/10.1016/j.wocn.2022.101174

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333-382.

Shaw, J. A., & Chen, W.-r. (2019). Spatially Conditioned Speech Timing: Evidence and Implications. *Frontiers in Psychology, 10*.

Shaw, J. A., & Gafos, A. I. (2015). Stochastic Time Models of Syllable Structure. *PLoS One, 10*, e0124714.

Shue, Y.-L., Keating, P., Vicenik, C., Yu, K. (2011) VoiceSauce: A program for voice analysis. In *Proceedings of ICPhS XVII*, 1846-1849.

Sorenson, T., Gafos, A. (2016). The Gesture as an Autonomous Nonlinear Dynamical System. *Ecological Psychology,* 28, 188-215.

Steffman, J., & Cole, J. (2022). An automated method for detecting F 0 measurement jumps based on sample-to-sample differences. *JASA Express Letters*, *2*(11), 115201.

Steffman, J., Shattuck-Hufnagel, S., & Cole, J. (2022). The rise and fall of American English pitch accents: Evidence from an imitation study of rising nuclear tunes. *Proc. Speech Prosody 2022*, 857-861.

Strogatz, S. H. (1994). Nonlinear Dynamics And Chaos: With Applications to Physics, Biology, Chemistry, And Engineering. Routledge.

't Hart, J., & Cohen, A. 1973. Intonation by rule: A perceptual quest. *Journal of Phonetics, 1*(4): 309-327. https://doi.org/10.1016/S0095-4470(19)31400-7.

Tilsen, S. (2019). Motoric Mechanisms for the Emergence of Non-local Phonological Patterns. *Frontiers in Psychology, 10*.

Trager, G. L., & Smith, H. L. (1951). *An Outline of English Structure*, Norman, OK: Battenburg Press. Reprinted 1957 by American Council of Learned Societies, Washington.

Turnbull, R. (2017). The role of predictability in intonational variability. L*anguage and Speech, 60*(1), 123–153.

Vanderslice, R., Ladefoged P. (1972). Binary Suprasegmental Features and Transformational Word-Accentuation Rules *Language*. 48: 819.

Veilleux, N., Shattuck-Hufnagel S. & Brugos A. *6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI.* January IAP 2006. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27, 55-105.